GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 11:13:02 ; Search time 17522 Seconds
(without alignments)
11673.789 Million cell updates/sec

Title: US-10-023-649-1_COPY_1_5000
Perfect score: 5000
Sequence: 1 accctacaaactaatcgatc.....ataattctagaaacgccacc 5000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5000	100.0	11484	14	AF214040	AF214040 Western e
2	3536	70.7	4471	14	WEU01065	U01065 Western equ
3	3536	70.7	4498	14	WEEVNS	X74892 Western equ
4	2766	55.3	11678	14	U01034	U01034 Eastern equ
5	2745.2	54.9	11675	14	EEEVIRNA	X63135 Eastern Equ
6	2133.4	42.7	11441	14	VEU55362	U55362 Venezuelan
7	2128.6	42.6	11420	14	AF004472	AF004472 Venezuela
8	2123.8	42.5	11420	14	AF004458	AF004458 Venezuela
9	2122.2	42.4	11444	14	EEVCOMGEN	L04653 Venezuelan
10	2120.6	42.4	11423	14	VEU55347	U55347 Venezuelan
11	2119.6	42.4	11447	14	EEVNSPEPA	L01442 Venezuelan
12	2119	42.4	11420	14	AF004459	AF004459 Venezuela
13	2119	42.4	11420	14	VEU55360	U55360 Venezuelan
14	2119	42.4	11423	14	AF375051	AF375051 Venezuela
15	2119	42.4	11423	14	VEU55342	U55342 Venezuelan
16	2118	42.4	11444	14	EEVNSPENV	J04332 Venezuelan
17	2118	42.4	11445	14	AF069903	AF069903 Venezuela
18	2118	42.4	11446	14	EEVNSPEPB	L01443 Venezuelan
19	2118	42.4	11459	6	AR170868	AR170868 Sequence
20	2117.4	42.3	11423	14	VEU55345	U55345 Venezuelan
21	2117.4	42.3	11423	14	VEU55350	U55350 Venezuelan
22	2115	42.3	11465	14	AF075257	AF075257 Venezuela
23	2114.8	42.3	11422	14	EEVNSPECFA	L00930 Venezuelan
24	2114.2	42.3	11421	14	AF100566	AF100566 Venezuela
25	2103.6	42.1	11391	14	AF075253	AF075253 Venezuela
26	2087.6	41.8	11464	14	AF448535	AF448535 Venezuela
27	2087.6	41.8	11464	14	AF448538	AF448538 Venezuela
28	2086.4	41.7	11385	14	AF075259	AF075259 Venezuela
29	2086	41.7	11461	14	AF448537	AF448537 Venezuela
30	2085.6	41.7	11344	14	AF075256	AF075256 Venezuela
31	2085.6	41.7	11464	14	AF448536	AF448536 Venezuela
32	2084.8	41.7	11464	14	AF448539	AF448539 Venezuela
33	2081.8	41.6	11395	14	AF075251	AF075251 Venezuela
34	2074.8	41.5	11464	14	VEU34999	U34999 Venezuelan
35	2070	41.4	11311	14	AF075255	AF075255 Venezuela
36	2049.2	41.0	11468	14	AF075252	AF075252 Venezuela
37	2039.4	40.8	11494	14	AF075258	AF075258 Venezuela
38	2031	40.6	11530	14	AF075254	AF075254 Venezuela
39	1676.8	33.5	11657	14	RRVNBCG	M20162 Ross River
40	1631.8	32.6	11698	14	AB032553	AB032553 Sagiyama
41	1589.2	31.8	14529	14	SFVRNAIS	Z48163 Semliki for
42	1584.8	31.7	7422	14	SFO251359	AJ251359 Semliki F
43	1573.8	31.5	11411	14	AF237947	AF237947 Mayaro vi
44	1567.4	31.3	11488	14	BFU73745	U73745 Barmah Fore
45	1567.4	31.3	11822	14	AF079456	AF079456 O'nyong-n

ALIGNMENTS

RESULT 1
AF214040
LOCUS AF214040 11484 bp RNA linear VRL 27-JAN-2000
DEFINITION Western equine encephalomyelitis virus strain 71V-1658 nonstructural polyprotein and structural polyprotein genes, complete cds.
ACCESSION AF214040 AF143811
VERSION AF214040.1 GI:6760410
KEYWORDS .
SOURCE Western equine encephalomyelitis virus
ORGANISM Western equine encephalomyelitis virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae; Alphavirus; WEEV complex.
REFERENCE 1 (bases 1 to 11484)
AUTHORS Netolitzky,D.J., Schmaltz,F.L., Parker,M.D., Rayner,G.A.,

```
          Fisher,G.R., Trent,D.W., Bader,D.E. and Nagata,L.P.
  TITLE   Complete genomic RNA sequence of western equine encephalitis virus
          and expression of the structural genes
  JOURNAL J. Gen. Virol. 81 (Pt 1), 151-159 (2000)
  MEDLINE 20109043
   PUBMED 10640553
REFERENCE 2  (bases 1 to 11484)
  AUTHORS Netolitzky,D.J., Schmaltz,F.L., Parker,M.D., Rayner,G.A.,
          Fisher,G.R., Trent,D.W., Bader,D.E. and Nagata,L.P.
  TITLE   Direct Submission
  JOURNAL Submitted (08-DEC-1999) Medical Countermeasures Section, Defence
          Research Establishment Suffield, P.O. Box 4000, Stn Main, Medicine
          Hat, Alberta T1A 8K6, Canada
FEATURES             Location/Qualifiers
     source          1. .11484
                     /organism="Western equine encephalomyelitis virus"
                     /mol_type="genomic RNA"
                     /strain="71V-1658"
                     /db_xref="taxon:11039"
     misc_feature    1. .11484
                     /note="sequence of this entry added to the sequence of
                     AF214039, strain CBA87, constitutes the Western equine
                     encephalomyelitis virus complete genome"
     5'UTR           <1. .24
     CDS             25. .7428
                     /note="encodes nsP1, nsP2, nsP3 and nsP4 proteins"
                     /codon_start=1
                     /transl_except=(pos:5581. .5583,aa:OTHER)
                     /product="nonstructural polyprotein"
                     /protein_id="AAF28339.1"
                     /db_xref="GI:6760411"
                     /translation="MERIHVDLDADSPYVKSLQRTFPQFEIEARQVTDNDHANARAFS
                     HVATKLIESEVDRDQVILDIGSAPVRHAHSNHRYHCICPMISAEDPDRLQRYAERLKK
                     SDITDKNIASKAADLLEVMSTPDAETPSLCMHTDATCRYFGSVAVYQDVYAVHAPTSI
                     YHQALKGVRTIYWIGFDTTPFMYKNMAGSYPTYNTNWADERVLEARNIGLGNSDLQES
                     RLGKLSILRKKRLQPTNKIIFSVGSTIYTEDRSLLRSWHLPNVFHLKGKSNFTGRCGT
                     IVSCEGYVIKKITISPGLYGKVENLASTMHREGFLSCKVTDTLRGERVSFAVCTYVPA
                     TLCDQMTGILATDVSVDDAQKLLVGLNQRIVVNGRTQRNTNTMQNYLLPVVAQAFSRW
                     AREHRADLDDEKELGVRERTLTMGCCWAFKTQKITSIYKKPGTQTIKKVPAVFDSFVI
                     PRLTSHGLDMGFRRRLKLLLEPTVKPAPAITMADVEHLRGLQQEAEEVAAAEEIREAL
                     PPLLPEIEKETVEAEVDLIMQEAGAGSVETPRGHIRVTSYPGEEKIGSYAILSPQAVL
                     NSEKLACIHPLAEQVLVMTHKGRAGRYKVEPYHGKVIVPEGTAVPVQDFQALSESATI
                     VFNEREFVNRYLHHIAINGGALNTDEEYYKTVKTQDTDSEYVFDIDARKCVKREDAGP
                     LCLTGDLVDPPFHEFAYESLKTRPAAPHKVPTIGVYGVPGSGKSGIIKSAVTKKDLVV
                     SAKKENCAEIIRDVRRMRRMDVAARTVDSVLLNGVKHPVNTLYIDEAFACHAGTLLAL
                     IAIVKPKKVVLCGDPKQCGFFNMMCLKVHFNHDICTEVYHKSISRRCTQTVTAIVSTL
                     FYDKRMKTVNPCADKIIIDTTGTTKPHKDDLILTCFRGWVKQLQIDYKNHEIMTAAAS
                     QGLTRKGVYAVRYKVNENPLYSQTSEHVNVLLTRTEKRIVWKTLAGDPWIKTLTAKYP
                     GDFTASLDDWQREHDAIMARVLDKPQTADVFQNKVNVCWAKALEPVLATANIVLTRQQ
                     WETLHPFKHDRAYSPEMALNFFCTRPFGVDLDSGLFSAPTVALTYRDQHWDNSPGKNM
                     YGLNREVAKELSRRYPCITKAVDTGRVADIRNNTIKDYSPTINVVPLNRRLPHSLIVD
                     HKGQGTTDHSGFLSKMKGKSVLVIGDPISIPGKKVESMGPLPTNTIRCDLDLGIPSHV
                     GKYDIIFVNVRTPYRNHHYQQCEDHAIHHSMLTCKAVHHLNTGGTCVAIGYGLADRAT
                     ENIITAVARSFRFTRVCQPKNTAENTEVLFVFFGKDNGNHTHDQDRLGVVLDNIYQGS
                     TRYEAGRAPAYRVIRGDISKSADQAIVNAANSKGQPGSGVCGALYRKWPAAFDRQPIA
                     VGTARLVKHEPLIIHAVGPNFSKMPEPEGDLKLAAAYMSIASIVNAERITKISVPLLS
                     TGIYSGGKDRVMQSLHHLFTAFDTTDADVTIYCLDKQWETRIIEAIHRKESVEILDDD
                     KPVDIDLVRVHPNSSLAGRPGYSVNEGKLYSYLEGTRFHQTAKDIAEIHAMWPNKSEA
                     NEQICLYILGESMSSIRSKCPVEESEASAPPHTLPCLCNYAMTAERVYRLRSAKKEQF
                     AVCSSFLLPKYRITGVQKLQCSKPVLFSGVVPPAVHPRKYAEIILETPPPPATTTVIC
                     EPTVPERIPSPVISRAPSAESLLSLGGVSFSSSATRSSTAWSDYDRRFVVTADVHQAN
                     TSTWSIPSAPGLDVQLPSDVTDSHWSIPSASGFEVRTPSVQDLTAECAKPRGLAEIMQ
                     DFNTAPFQFLSDYRPVPAPRRRPIPSPRSTASAPPVPKPRRTKYQQPPGVARAISEAE
                     LDEYIRQHSNXRYEAGAYIFSSETGQGHLQQKSVRQCKLQEPILDRAVHEKYYAPRLD
                     LEREKMLQKKLQLCASEGNRSRYQSRKVENMKAITAERLISGLGTYLSSEVNPVECYR
                     VNYPVPIYSSTVINRFTSAEVAVKTCNLVIQENYPTVASYCITDEYDAYLDMVDGASC
                     CLDTATFCPAKLRSYPKKHSYLQPEIRSAVPSPIQNTLQNVLAAATKRNCNVTQMREL
                     PVLDSAAFNVDCFKKYACNDEYWDTFRDNPIRLTTENVTQYVTKLKGPKAAALFANTH
                     NLKPLQEIPMDQFVMDLKRDVKVTPGTKHTEERPKVQVIQAADPLATAYLCGIHRELV
                     RRLNAVLLPNIHTLFDMSAEDFDAIIAEHFHHGDPVLETDIASFDKSEDDAIAISALM
                     ILEDLGVDQPLLDLIEAAFGNITSVHLPTGTRFKFGAMMKSGMFLTLFVNTLVNIMIA
                     SRVLRERLTTSACAASIGDDNIVHGVVSDTLMAERCATWLNMEVKIIDAVIGIKAPYF
                     CGGFILVDQITGTACRVADPLKRLFKLGKPLPVDDTQDCDRRRALHDEAMRWNRIGIT
                     DELVKAVESRYEIILAGLIITSLSTLAESVKNFKSIRGSPITLYG"
     CDS             7473. .11183
                     /note="encodes nucleocapsid, E3, E2, 6K and E1 proteins"
                     /codon_start=1
                     /product="structural polyprotein"
                     /protein_id="AAF28340.1"
                     /db_xref="GI:6760412"
                     /translation="MFPYPQLNFPPVYPTNPMAYRDPNPPRRRWRPFRPPLAAQIEDL
                     RRSIVNLTFKQRSPNPPPGPPPKKKKSAPKPKPTQPKKKKQQAKRTKRKPKPGKRQRM
                     CMKLESDKTFPIMLNGQVNGYACVVGGRLMKPLHVEGKIDNEQLAAVKLKKASMYDLE
                     YGDVPQNMKSDTLQYTSDKPPGFYNWHHGAVQYENGRFTVPRGVGGKGDSGRPILDNR
                     GRVVAIVLGGANEGTRTALSVVTWNQKGVTIRDTPEGSEPWSLVTALCVLSNVTFPCD
                     KPPVCYSLTPERTLDVLEENVDNPNYDTLLENVLKCPSRRPKRSITDDFTLTSPYLGF
                     CPYCRHSTPCFSPIKIENVWDESDDGSIRIQVSAQFGYNQAGTADVTKFRYMSFDHDH
                     DIKEDSMEKIAISTSGPCRRLGHKGYFLLAQCPPGDSVTVSITSGASENSCTVEKKIR
                     RKFVGREEYLFPPVHGKLVKCHVYDHLKETSAGYITMHRPGPHAYKSYLEEASGEVYI
                     KPPSGKNVTYECKCGDYSTGIVSTRTKMNGCTKAKQCIAYKSDQTKWVFNSPDLIRHT
                     DHSVQGKLHIPFRLTPTVCPVPLAHTPTVTKWFKGITLHLTAMRPTLLTTRKLGLRAD
                     ATAEWITGSTSRNFSVGREGLEYVWGNHEPVRVWAQESAPGDPHGWPHEIIIHYYHRH
                     PVYTVIVLCGVALAILVGTASSAACIAKARRDCLTPYALAPNATVPTALAVLCCIRPT
                     NAETFGETLNHLWFNNQPFLWAQLCIPLAALVILFRCFSCCMPFLLVAGVCLGKVDAF
                     EHATTVPNVPGIPYKALVERAGYAPLNLEITVVSSELTPSTNKEYVTCKFHTVIPSPQ
                     VKCCGSLECKASSKADYTCRVFGGVYPFMWGGAQCFCDSENTQLSEAYVEFAPDCTID
                     HAVALKVHTAALKVGLRIVYGNTTAHLDTFVNGVTPGSSRDLKVIAGPISAAFSPFDH
                     KVVIRKGLVYNYDFPEYGAMKPGAFGDIQASSLDATDIVARTDIRLLKPSVKNIHVPY
                     TQAVSGYEMWKNNSGRPLQETAPFGCKIEVEPLRASNCAYGHIPISIDIPDAAFVRSS
                     ESPTILEVSCTVADCIYSADFGGSLTLQYKADREGHCPVHSHSTTAVLKEATTHVTAV
                     GSITLHFSTSSPQANFIVSLCGKKSTCNAECKPPADHIIGEPHKVDQEFQAAVSKTSW
                     NWLLALFGGASSLIVVGLIVLVCSSMLINTRR"
     3'UTR           11184. .11484
BASE COUNT     3251 a    2842 c    2808 g    2583 t
ORIGIN
```

```
  Query Match             100.0%;  Score 5000;  DB 14;  Length 11484;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 5000;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ACCCTACAAACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ACCCTACAAACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGC 60

Qy         61 CCGTATGTCAAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGTATGTCAAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTC 120

Qy        121 ACTGACAATGACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACTGACAATGACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAG 180

Qy        181 AGCGAAGTCGACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGAAGTCGACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCA 240

Qy        241 CATTCCAATCACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATTCCAATCACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGA 300

Qy        301 CTACAACGGTATGCAGAAAGACTTAAGAAAAGTGACATTACCGACAAGAACATAGCCTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTACAACGGTATGCAGAAAGACTTAAGAAAAGTGACATTACCGACAAGAACATAGCCTCT 360

Qy        361 AAGGCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAGGCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGT 420

Qy        421 ATGCACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTAC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ATGCACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTAC 480

Qy        481 GCAGTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCAGTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTAC 540
```

```
Qy       541 TGGATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACT 600
Db       541 TGGATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACT 600

Qy       601 TACAACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAAC 660
Db       601 TACAACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAAC 660

Qy       661 TCAGATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAA 720
Db       661 TCAGATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAA 720

Qy       721 CCTACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTG 780
Db       721 CCTACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTG 780

Qy       781 TTACGTAGCTGGCATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGT 840
Db       781 TTACGTAGCTGGCATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGT 840

Qy       841 AGATGTGGGACCATTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCA 900
Db       841 AGATGTGGGACCATTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCA 900

Qy       901 GGACTATACGGTAAAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGT 960
Db       901 GGACTATACGGTAAAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGT 960

Qy       961 TGCAAAGTCACAGATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTA 1020
Db       961 TGCAAAGTCACAGATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTA 1020

Qy      1021 CCAGCCACACTTTGCGATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGAC 1080
Db      1021 CCAGCCACACTTTGCGATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGAC 1080

Qy      1081 GCACAAAAACTATTGGTTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGA 1140
Db      1081 GCACAAAAACTATTGGTTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGA 1140

Qy      1141 AATACTAACACAATGCAGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGG 1200
Db      1141 AATACTAACACAATGCAGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGG 1200

Qy      1201 GCGCGTGAACATCGTGCCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACT 1260
Db      1201 GCGCGTGAACATCGTGCCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACT 1260

Qy      1261 CTTACTATGGGCTGCTGCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAGAAG 1320
Db      1261 CTTACTATGGGCTGCTGCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAGAAG 1320

Qy      1321 CCTGGTACGCAAACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGC 1380
Db      1321 CCTGGTACGCAAACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGC 1380

Qy      1381 CTTACCAGCCACGGGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCA 1440
Db      1381 CTTACCAGCCACGGGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCA 1440

Qy      1441 ACTGTCAAACCCGCACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAG 1500
Db      1441 ACTGTCAAACCCGCACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAG 1500

Qy      1501 CAAGAAGCTGAAGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTC 1560
Db      1501 CAAGAAGCTGAAGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTC 1560

Qy      1561 CCTGAAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGA 1620
Db      1561 CCTGAAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGA 1620

Qy      1621 GCAGGTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAG 1680
```

```
Db      1621 GCAGGTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAG 1680

Qy      1681 AAGATTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCG 1740
Db      1681 AAGATTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCG 1740

Qy      1741 TGTATCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGA 1800
Db      1741 TGTATCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGA 1800

Qy      1801 TACAAAGTCGAGCCATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTT 1860
Db      1801 TACAAAGTCGAGCCATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTT 1860

Qy      1861 CAAGACTTCCAGGCATTGAGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTA 1920
Db      1861 CAAGACTTCCAGGCATTGAGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTA 1920

Qy      1921 AACAGATACCTGCACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTAC 1980
Db      1921 AACAGATACCTGCACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTAC 1980

Qy      1981 TATAAGACTGTAAAGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGA 2040
Db      1981 TATAAGACTGTAAAGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGA 2040

Qy      2041 AAGTGTGTTAAGCGAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCA 2100
Db      2041 AAGTGTGTTAAGCGAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCA 2100

Qy      2101 CCATTTCACGAGTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTC 2160
Db      2101 CCATTTCACGAGTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTC 2160

Qy      2161 CCAACCATCGGAGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCT 2220
Db      2161 CCAACCATCGGAGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCT 2220

Qy      2221 GTGACTAAGAAAGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGG 2280
Db      2221 GTGACTAAGAAAGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGG 2280

Qy      2281 GATGTAAGGAGGATGAGACGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTA 2340
Db      2281 GATGTAAGGAGGATGAGACGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTA 2340

Qy      2341 AATGGGGTTAAGCACCCCGTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCA 2400
Db      2341 AATGGGGTTAAGCACCCCGTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCA 2400

Qy      2401 GGGACGCTGCTGGCACTGATTGCCATCGTCAAACCTAAGAAAGTGGTATTGTGCGGGGAC 2460
Db      2401 GGGACGCTGCTGGCACTGATTGCCATCGTCAAACCTAAGAAAGTGGTATTGTGCGGGGAC 2460

Qy      2461 CCAAAACAATGCGGCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATA 2520
Db      2461 CCAAAACAATGCGGCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATA 2520

Qy      2521 TGCACTGAAGTGTACCATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATC 2580
Db      2521 TGCACTGAAGTGTACCATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATC 2580

Qy      2581 GTCTCCACGCTCTTCTACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATC 2640
Db      2581 GTCTCCACGCTCTTCTACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATC 2640

Qy      2641 ATCATAGATACCACAGGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTC 2700
Db      2641 ATCATAGATACCACAGGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTC 2700

Qy      2701 AGAGGATGGGTGAAACAGCTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCT 2760
```

```
Db      2701 AGAGGATGGGTGAAACAGCTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCT 2760
Qy      2761 GCATCGCAAGGACTTACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAAT 2820
Db      2761 GCATCGCAAGGACTTACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAAT 2820
Qy      2821 CCACTCTACTCGCAGACTTCTGAGCACGTGAACGTGTTACTTACACGCACAGAAAAACGC 2880
Db      2821 CCACTCTACTCGCAGACTTCTGAGCACGTGAACGTGTTACTTACACGCACAGAAAAACGC 2880
Qy      2881 ATTGTCTGGAAGACGCTAGCTGGTGATCCCTGGATAAAGACACTTACAGCTAAATATCCC 2940
Db      2881 ATTGTCTGGAAGACGCTAGCTGGTGATCCCTGGATAAAGACACTTACAGCTAAATATCCC 2940
Qy      2941 GGGGATTTCACGGCTTCATTGGACGACTGGCAGCGCGAACACGACGCCATTATGGCACGC 3000
Db      2941 GGGGATTTCACGGCTTCATTGGACGACTGGCAGCGCGAACACGACGCCATTATGGCACGC 3000
Qy      3001 GTTCTTGATAAGCCGCAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCG 3060
Db      3001 GTTCTTGATAAGCCGCAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCG 3060
Qy      3061 AAGGCTTTAGAGCCAGTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAG 3120
Db      3061 AAGGCTTTAGAGCCAGTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAG 3120
Qy      3121 ACGTTGCACCCATTCAAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTT 3180
Db      3121 ACGTTGCACCCATTCAAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTT 3180
Qy      3181 TGCACCAGGTTCTTTGGAGTAGACCTGGACAGTGGGTTATTTTCCGCTCCTACCGTCGCA 3240
Db      3181 TGCACCAGGTTCTTTGGAGTAGACCTGGACAGTGGGTTATTTTCCGCTCCTACCGTCGCA 3240
Qy      3241 CTTACTTACAGGGATCAGCACTGGGATAACTCGCCAGGGAAGAACATGTATGGGCTTAAT 3300
Db      3241 CTTACTTACAGGGATCAGCACTGGGATAACTCGCCAGGGAAGAACATGTATGGGCTTAAT 3300
Qy      3301 AGAGAGGTAGCAAAGGAGTTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACA 3360
Db      3301 AGAGAGGTAGCAAAGGAGTTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACA 3360
Qy      3361 GGCAGGGTAGCTGATATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTG 3420
Db      3361 GGCAGGGTAGCTGATATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTG 3420
Qy      3421 GTTCCATTAAATCGCCGGTTGCCCCACTCGTTGATCGTTGACCACAAAGGACAGGGTACA 3480
Db      3421 GTTCCATTAAATCGCCGGTTGCCCCACTCGTTGATCGTTGACCACAAAGGACAGGGTACA 3480
Qy      3481 ACTGATCACAGCGGATTCCTATCTAAGATGAAGGGCAAATCTGTGTTGGTGATCGGCGAT 3540
Db      3481 ACTGATCACAGCGGATTCCTATCTAAGATGAAGGGCAAATCTGTGTTGGTGATCGGCGAT 3540
Qy      3541 CCTATCAGCATTCCAGGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATACCATC 3600
Db      3541 CCTATCAGCATTCCAGGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATACCATC 3600
Qy      3601 AGGTGTGATCTCGATTTGGGAATACCTAGCCATGTCGGTAAATATGACATTATCTTTGTC 3660
Db      3601 AGGTGTGATCTCGATTTGGGAATACCTAGCCATGTCGGTAAATATGACATTATCTTTGTC 3660
Qy      3661 AATGTTAGGACCCCGTACAGGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCAC 3720
Db      3661 AATGTTAGGACCCCGTACAGGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCAC 3720
Qy      3721 CACAGCATGCTAACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCT 3780
Db      3721 CACAGCATGCTAACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCT 3780
Qy      3781 ATAGGGTATGGGCTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCACGCTCA 3840
Db      3781 ATAGGGTATGGGCTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCACGCTCA 3840
```

```
Qy      3841 TTTAGGTTTACCCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTC 3900
Db      3841 TTTAGGTTTACCCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTC 3900
Qy      3901 GTGTTCTTCGGCAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTG 3960
Db      3901 GTGTTCTTCGGCAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTG 3960
Qy      3961 CTTGACAACATCTATCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGTACAGA 4020
Db      3961 CTTGACAACATCTATCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGTACAGA 4020
Qy      4021 GTGATCAGAGGTGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGC 4080
Db      4021 GTGATCAGAGGTGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGC 4080
Qy      4081 AAAGGTCAACCAGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTT 4140
Db      4081 AAAGGTCAACCAGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTT 4140
Qy      4141 GATAGACAGCCAATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACCGCTCATCATA 4200
Db      4141 GATAGACAGCCAATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACCGCTCATCATA 4200
Qy      4201 CATGCTGTAGGACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTAAGCTCGCA 4260
Db      4201 CATGCTGTAGGACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTAAGCTCGCA 4260
Qy      4261 GCTGCCTACATGAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTA 4320
Db      4261 GCTGCCTACATGAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTA 4320
Qy      4321 CCGCTACTGTCAACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCAT 4380
Db      4321 CCGCTACTGTCAACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCAT 4380
Qy      4381 CACCTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAA 4440
Db      4381 CACCTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAA 4440
Qy      4441 CAATGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATTCTGGAT 4500
Db      4441 CAATGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATTCTGGAT 4500
Qy      4501 GATGACAAGCCAGTAGACATTGACTTGGTCAGGGTCCACCCAAACAGCTCTTTGGCAGGC 4560
Db      4501 GATGACAAGCCAGTAGACATTGACTTGGTCAGGGTCCACCCAAACAGCTCTTTGGCAGGC 4560
Qy      4561 AGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTGGAAGGTACACGATTC 4620
Db      4561 AGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTGGAAGGTACACGATTC 4620
Qy      4621 CATCAGACCGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCCAACAAATCTGAGGCT 4680
Db      4621 CATCAGACCGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCCAACAAATCTGAGGCT 4680
Qy      4681 AATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGC 4740
Db      4681 AATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGC 4740
Qy      4741 CCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTAC 4800
Db      4741 CCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTAC 4800
Qy      4801 GCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTA 4860
Db      4801 GCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTA 4860
Qy      4861 TGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGC 4920
Db      4861 TGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGC 4920
```

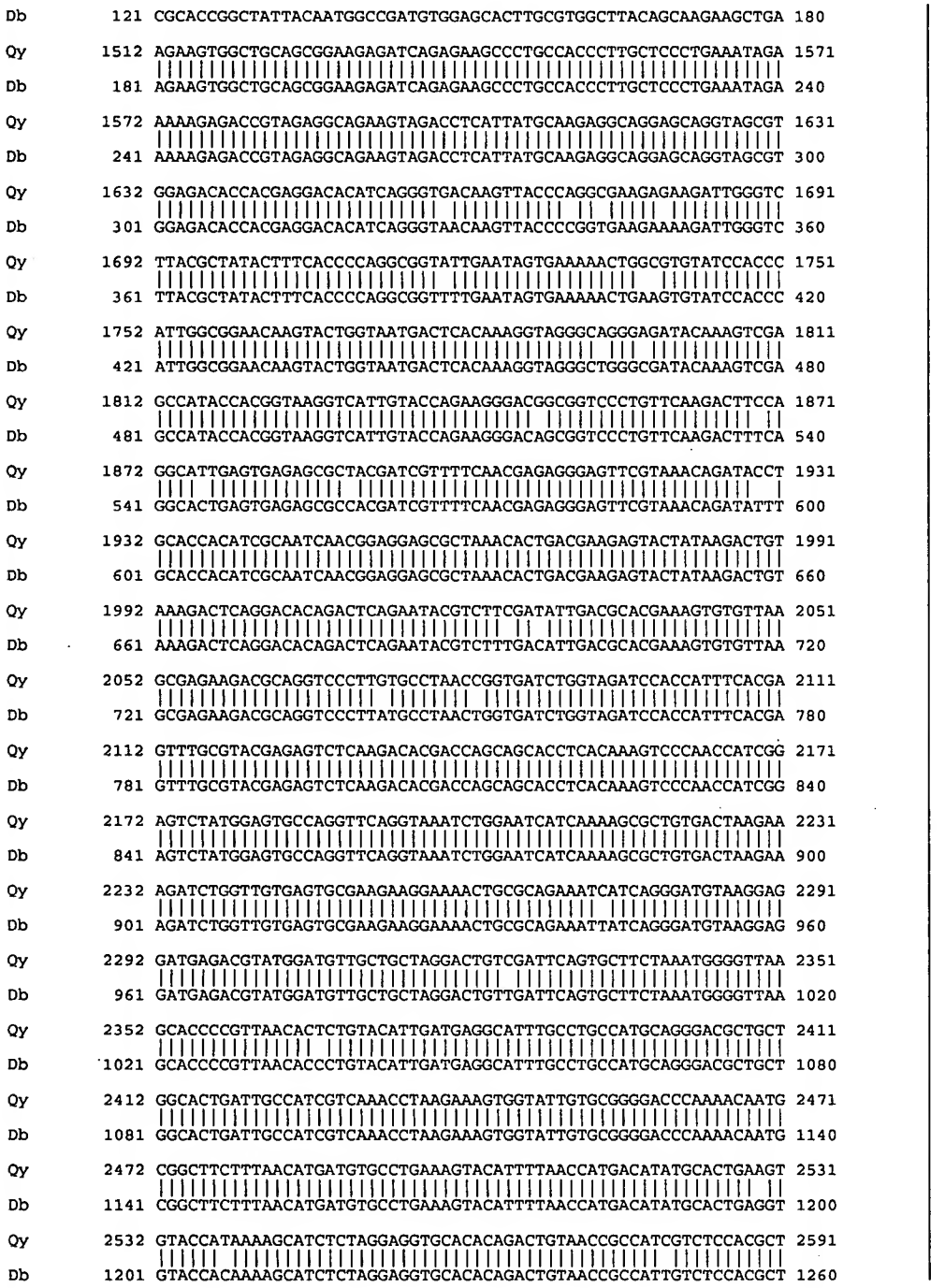
```
Qy       4921 AAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGCGGAA 4980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4921 AAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGCGGAA 4980

Qy       4981 ATAATTCTAGAAACGCCACC 5000
             ||||||||||||||||||||
Db       4981 ATAATTCTAGAAACGCCACC 5000


RESULT 2
WEU01065
LOCUS       WEU01065                4471 bp    RNA     linear   VRL 01-JUN-1995
DEFINITION  Western equine encephalomyelitis virus nonstructural polyprotein
            gene, partial cds.
ACCESSION   U01065
VERSION     U01065.1  GI:393033
KEYWORDS    nsP1; nsP2; nsP3; nsP4.
SOURCE      Western equine encephalomyelitis virus
  ORGANISM  Western equine encephalomyelitis virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
            Alphavirus; WEEV complex.
REFERENCE   1  (bases 1 to 4471)
  AUTHORS   Weaver,S.C., Hagenbaugh,A., Bellew,L., Netesov,S.V., Volchkov,V.E.,
            Chang,G.J., Clarke,D.K., Gousset,L., Scott,T.W., Trent,D.W. and
            Holland,J.J.
  TITLE     A comparison of the nucleotide sequences of eastern and western
            equine encephalomyelitis viruses with those of other alphaviruses
            and related RNA viruses
  JOURNAL   Virology 197 (1), 375-390 (1993)
  MEDLINE   94025587
   PUBMED   8105605
REFERENCE   2  (bases 1 to 4471)
  AUTHORS   Weaver,S.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-1993) Scott C. Weaver, Biology, University of
            California, San Diego, La Jolla, CA 92037-0116 USA
FEATURES             Location/Qualifiers
     source          1. .4471
                     /organism="Western equine encephalomyelitis virus"
                     /mol_type="genomic RNA"
                     /isolate="5614"
                     /db_xref="taxon:11039"
                     /clone="83, 34, 44, 75, 312"
                     /note="Western equine encephalitis virus"
     CDS             <1. .>4471
                     /note="termination codon at position 4253. .4255"
                     /codon_start=2
                     /transl_except=(pos:4253. .4255,aa:OTHER)
                     /product="nonstructural polyprotein"
                     /protein_id="AAA67651.1"
                     /db_xref="GI:393034"
                     /translation="TIKKVPAVFDSFVIPRLTSHGPDIGLPRRSRLLLEPTVKPAPAI
                     TMADVEHLRGLQQEAEEVAAAEEIREALPPLLPEIEKETVEAEVDLIMQEAGAGSVET
                     PRGHIRVTSYPGEEKIGSYAILSPQAVLNSEKLKCIHPLAEQVLVMTHKGRAGRYKVE
                     PYHGKVIVPEGTAVPVQDFQALSESATIVFNEREFVNRYLHHIAINGGALNTDEEYYK
                     TVKTQDTDSEYVFDIDARKCVKREDAGPLCLTGDLVDPPFHEFAYESLKTRPAAPHKV
                     PTIGVYGVPGSGKSGIIKSAVTKKDLVVSAKKENCAEIIRDVRRMRRMDVAARTVDSV
                     LLNGVKHPVNTLYIDEAFACHAGTLLALIAIVKPKKVVLCGDPKQCGFFNMMCLKVHF
                     NHDICTEVYHKSISRRCTQTVTAIVSTLFYDKRMKTVNPCADKIIIDTTGTTKPHKDD
                     LILTCFRGWVKQLQIDYKNHEIMTAAASQGLTRKGVYAVRYKVNENPLYSQTSEHVNV
                     LVTRTEKRIVWKTLAGDPWIKTFTAKYPGDFTASLDDWQREHDAIMARVLDKPQTADV
                     FQNKVNVCWAKALEPVLATANIVLTRQQWETLHPFKHDRAYSPEMALNFFCTRFFGVD
                     LDSGLFSAPTVALTYRDQHWDNSPGKNMYGLNREVAKELSRRYPCITKAVDTGRVADI
                     RNNTIKDYSPTINVVPLNRRLPHSLIVDHKGQGTTDHSGFLSKMNGKSVLVIGDPISI
                     PGKKVESMGPLPTNTIRCDLDLGIPSHVGKYDIIFVNVRTPYKNHHYQQCEDHAIHHS
                     MLTCKAVHHLNTGGTCVAIGYGLADRATENIITAVARSFRFTRVCQPKNTAENTEVLF
                     VFFGKDNGNHTHDQDRLGVVLDNIYQGSTRYEAGRAPAYRVIRGDISKSADQAIVNAA
                     NSKGQPGSGVCGALYRKWPAAFDRQPIAVGTARLVKHEPLIIHAVGPNFSKMPEREGD
                     LKLAAAYMSIASIVNAERITKISVPLLSTGIYSGGKDRVMQSLHHLFTAFDTTDADVT
                     IYCLDKQWETRITEAIHRKESVEILDDDDKPVDIDLVRVHPNSSLAGRPGYSVNEGKL
                     YSYLEGTRFHQTAKDIAEIHAMWPNKSEANEQICLYILGESMSSIRSKCPVEESEASA
                     PPHTLPCLCNYAMTAERVYRLRSAKKEQFAVCSSFPLPKYRITGVQKLQCSKPVLFSG
                     VVPPAVHPRKYAEIILETPPSPTTTTVICEPTVPERIPSPVISRAPSAESLLSFGGVS
                     FSSSATRSSTAWSDYDRRFVVTADVHQANTSTWSIPSAPGLDVQLPSDDTDSHWSIPS
                     ALGFEVRTPSVQDLTAECARPRGLAEIMQDFNTAPFQFLSDHRPVPAPRRRPIPSPRS
                     TASAPPVPKPRRTKYQQPPGVARAISEAELDEYIRQHSNXRYEAGAYIFSSETGQGHL
                     QQKSVRQCKLQDPIFERAVHEKYYAPRLDLEREKMLQKKLQLCASEGNRSRYQS"
     mat_peptide     <1. .292
                     /product="nsP1 protein"
     mat_peptide     293. .2674
                     /product="nsP2 protein"
     mat_peptide     2675. .4273
                     /product="nsP3 protein"
     mat_peptide     4274. .>4471
                     /product="nsP4 protein"
BASE COUNT     1272 a   1110 c   1116 g    973 t
ORIGIN

  Query Match           70.7%;  Score 3536;  DB 14;  Length 4471;
  Best Local Similarity   97.9%;  Pred. No. 0;
  Matches 3594;  Conservative    0;  Mismatches   75;  Indels    3;  Gaps    1;

Qy       1332 AACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCA 1391
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCA 60

Qy       1392 CGGGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACTGTCAAACC 1451
             |||||  ||||| ||  | |  ||||||  || ||||||||||||||||||||||||||||
Db         61 CGGGCCTGATATAGGGCTTCCGCGTAGGTCCAGGCTGCTGCTTGAACCAACTGTCAAACC 120

Qy       1452 CGCACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGA 1511
             ||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db         121 CGCACCGGCTATTACAATGGCCGATGTGGAGCACTTGCGTGGCTTACAGCAAGAAGCTGA 180

Qy       1512 AGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGA 1571
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 AGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGA 240

Qy       1572 AAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGAGCAGGTAGCGT 1631
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 AAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGAGCAGGTAGCGT 300

Qy       1632 GGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAGATTGGGTC 1691
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 GGAGACACCACGAGGACACATCAGGGTAACAAGTTACCCCGGTGAAGAAAAGATTGGGTC 360

Qy       1692 TTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGTATCCACCC 1751
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 TTACGCTATACTTTCACCCCAGGCGGTTTTGAATAGTGAAAAACTGAAGTGTATCCACCC 420

Qy       1752 ATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATACAAAGTCGA 1811
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 ATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCTGGGCGATACAAAGTCGA 480

Qy       1812 GCCATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAAGACTTCCA 1871
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 GCCATACCACGGTAAGGTCATTGTACCAGAAGGGACAGCGGTCCCTGTTCAAGACTTTCA 540

Qy       1872 GGCATTGAGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAACAGATACCT 1931
             |||| ||||||||||||| ||||||||||||||||||||||||||||||||||||||||  |
Db         541 GGCACTGAGTGAGAGCGCCACGATCGTTTTCAACGAGAGGGAGTTCGTAAACAGATATTT 600

Qy       1932 GCACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTATAAGACTGT 1991
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 GCACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTATAAGACTGT 660

Qy       1992 AAAGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAGTGTGTTAA 2051
             ||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db         661 AAAGACTCAGGACACAGACTCAGAATACGTCTTTGACATTGACGCACGAAAGTGTGTTAA 720

Qy       2052 GCGAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCATTTCACGA 2111
             ||||||||||||||||||||| ||||||||  ||||||||||||||||||||||||||||||
Db         721 GCGAGAAGACGCAGGTCCCTTATGCCTAACTGGTGATCTGGTAGATCCACCATTTCACGA 780

Qy       2112 GTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCAACCATCGG 2171
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        781 GTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCAACCATCGG 840
Qy       2172 AGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTGACTAAGAA 2231
Db        841 AGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTGACTAAGAA 900
Qy       2232 AGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGATGTAAGGAG 2291
Db        901 AGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATTATCAGGGATGTAAGGAG 960
Qy       2292 GATGAGACGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAATGGGGTTAA 2351
Db        961 GATGAGACGTATGGATGTTGCTGCTAGGACTGTTGATTCAGTGCTTCTAAATGGGGTTAA 1020
Qy       2352 GCACCCCGTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGGACGCTGCT 2411
Db       1021 GCACCCCGTTAACACCCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGGACGCTGCT 1080
Qy       2412 GGCACTGATTGCCATCGTCAAACCTAAGAAAGTGGTATTGTGCGGGGACCCAAAACAATG 2471
Db       1081 GGCACTGATTGCCATCGTCAAACCTAAGAAAGTGGTATTGTGCGGGGACCCAAAACAATG 1140
Qy       2472 CGGCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGCACTGAAGT 2531
Db       1141 CGGCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGCACTGAGGT 1200
Qy       2532 GTACCATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATCGTCTCCACGCT 2591
Db       1201 GTACCACAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATTGTCTCCACGCT 1260
Qy       2592 CTTCTACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATCATCATAGATAC 2651
Db       1261 CTTCTACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGACAAAATCATCATAGATAC 1320
Qy       2652 CACAGGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGAGGATGGGT 2711
Db       1321 CACAGGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGAGGATGGGT 1380
Qy       2712 GAAACAGCTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCTGCATCGCAAGG 2771
Db       1381 GAAACAGCTACAGATTGACTATAAAAATCATGAAATCATGACTGCGGCTGCATCGCAAGG 1440
Qy       2772 ACTTACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCACTCTACTC 2831
Db       1441 ACTTACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCACTCTACTC 1500
Qy       2832 GCAGACTTCTGAGCACGTGAACGTGTTACTTACACGCACAGAAAAACGCATTGTCTGGAA 2891
Db       1501 GCAGACTTCTGAGCACGTGAACGTGTTAGTTACACGCACAGAAAAACGCATTGTCTGGAA 1560
Qy       2892 GACGCTAGCTGGTGATCCCTGGATAAAGACACTTACAGCTAAATATCCCGGGGATTTCAC 2951
Db       1561 GACGCTAGCTGGTGACCCCTGGATAAAGACATTTACAGCTAAATACCCCGGGGATTTCAC 1620
Qy       2952 GGCTTCATTGGACGACTGGCAGCGCGAACACGACGCCATTATGGCACGCGTTCTTGATAA 3011
Db       1621 GGCTTCATTGGACGACTGGCAGCGCGAACACGACGCTATTATGGCACGCGTTCTTGATAA 1680
Qy       3012 GCCGCAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAGGCTTTAGA 3071
Db       1681 ACCGCAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAGGCTTTAGA 1740
Qy       3072 GCCAGTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAGACGTTGCACCC 3131
Db       1741 GCCAGTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAGACGTTGCACCC 1800
Qy       3132 ATTCAAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTTTGCACCAGGTT 3191
Db       1801 ATTCAAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTTTGCACCAGGTT 1860
Qy       3192 CTTTGGAGTAGACCTGGACAGTGGGTTATTTTCCGCTCCTACCGTCGCACTTACTTACAG 3251
Db       1861 CTTTGGTGTAGACCTGGACAGTGGGTTGTTTTCCGCTCCTACCGTCGCACTTACTTACAG 1920

Qy       3252 GGATCAGCACTGGGATAACTCGCCAGGGAAGAACATGTATGGGCTTAATAGAGAGGTAGC 3311
Db       1921 GGATCAGCACTGGGACAACTCGCCAGGGAAGAACATGTATGGGCTTAATAGAGAGGTAGC 1980
Qy       3312 AAAGGAGTTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACAGGCAGGGTAGC 3371
Db       1981 AAAGGAGCTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACAGGCAGGGTAGC 2040
Qy       3372 TGATATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTGGTTCCATTAAA 3431
Db       2041 TGATATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTGGTTCCATTAAA 2100
Qy       3432 TCGCCGGTTGCCCCACTCGTTGATCGTTGACCACAAAGGACAGGGTACAACTGATCACAG 3491
Db       2101 TCGCCGGTTGCCCCACTCGCTGATCGTTGACCACAAAGGACAGGGTACAACTGATCACAG 2160
Qy       3492 CGGATTCCTATCTAAGATGAAGGGCAAATCTGTGTTGGTGATCGGCGATCCTATCAGCAT 3551
Db       2161 TGGATTCCTATCTAAGATGAATGGCAAATCTGTGTTGGTGATCGGCGATCCTATCAGCAT 2220
Qy       3552 TCCAGGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATACCATCAGGTGTGATCT 3611
Db       2221 TCCTGGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATACCATCAGGTGTGATCT 2280
Qy       3612 CGATTTGGGAATACCTAGCCATGTCGGTAAATATGACATTATCTTTGTCAATGTTAGGAC 3671
Db       2281 CGATTTGGGAATACCTAGCCATGTCGGTAAATATGACATTATATTTGTCAATGTTAGGAC 2340
Qy       3672 CCCGTACAGGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCACCACAGCATGCT 3731
Db       2341 CCCATACAAGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCACCACAGCATGCT 2400
Qy       3732 AACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCTATAGGGTATGG 3791
Db       2401 AACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCCATAGGGTATGG 2460
Qy       3792 GCTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCACGCTCATTTAGGTTTAC 3851
Db       2461 ACTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCTCGCTCATTTAGGTTTAC 2520
Qy       3852 CCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTCGTGTTCTTCGG 3911
Db       2521 CCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTCGTGTTCTTCGG 2580
Qy       3912 CAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTGCTTGACAACAT 3971
Db       2581 CAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTGCTTGACAACAT 2640
Qy       3972 CTATCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGTACAGAGTGATCAGAGG 4031
Db       2641 CTACCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGTACAGAGTGATCAGAGG 2700
Qy       4032 TGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGCAAAGGTCAACC 4091
Db       2701 TGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGCAAAGGTCAACC 2760
Qy       4092 AGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTTGATAGACAGCC 4151
Db       2761 AGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTTGATAGACAGCC 2820
Qy       4152 AATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACCGCTCATCATACATGCTGTAGG 4211
Db       2821 AATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACCGCTCATCATACATGCTGTAGG 2880
Qy       4212 ACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTAAGCTCGCAGCTGCCTACAT 4271
Db       2881 ACCGAATTTTTCTAAGATGCCGGAACGAGAGGGCGACCTTAAGCTCGCAGCTGCCTACAT 2940
Qy       4272 GAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTACCGCTACTGTC 4331
Db       2941 GAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACTAAAATATCAGTACCGCTACTGTC 3000
```

```
Qy       4332 AACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCATCACCTGTTCAC 4391
             ||| ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       3001 AACTGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCATCACTTGTTCAC 3060

Qy       4392 TGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAACAATGGGAGAC 4451
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 TGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAACAATGGGAGAC 3120

Qy       4452 CAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATTCTG---GATGATGACAA 4508
             |||||||| ||||||||||||||||||||||||||||||||||| |||   ||||||||||
Db       3121 CAGGATAACCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATACTGGATGATGATGACAA 3180

Qy       4509 GCCAGTAGACATTGACTTGGTCAGGGTCCACCCAAACAGCTCTTTGGCAGGCAGACCAGG 4568
             |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       3181 GCCAGTAGACATCGACTTGGTCAGGGTCCACCCAAACAGCTCTTTGGCAGGCAGACCAGG 3240

Qy       4569 TTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTGGAAGGTACACGATTCCATCAGAC 4628
             |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       3241 TTACTCCGTCAATGAGGGCAAGCTGTATTCATACCTGGAAGGTACACGATTCCATCAGAC 3300

Qy       4629 CGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCCAACAAATCTGAGGCTAATGAGCA 4688
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 CGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCCAACAAATCTGAGGCTAATGAGCA 3360

Qy       4689 GATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGCCCAGTAGA 4748
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 GATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGCCCAGTAGA 3420

Qy       4749 GGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTACGCTATGAC 4808
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 GGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTACGCTATGAC 3480

Qy       4809 GGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTATGCTCATC 4868
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 GGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTATGCTCATC 3540

Qy       4869 ATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGCAAACCAGT 4928
             ||||| |||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       3541 ATTCCCGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAATGCAGCAAACCAGT 3600

Qy       4929 CCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGCGGAAATAATTCT 4988
             |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 CCTGTTCTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGCGGAAATAATTCT 3660

Qy       4989 AGAAACGCCACC 5000
             ||||||||||||
Db       3661 AGAAACGCCACC 3672
```

```
RESULT 3
WEEVNS
LOCUS       WEEVNS                  4498 bp    RNA     linear   VRL 14-FEB-2003
DEFINITION  Western equine encephalomyelitis virus partial gene for
            nonstructural polyprotein, genomic RNA.
ACCESSION   X74892
VERSION     X74892.1  GI:398206
KEYWORDS    nonstructural polyprotein; NSP1; NSP2; NSP3; NSP4.
SOURCE      Western equine encephalomyelitis virus
  ORGANISM  Western equine encephalomyelitis virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
            Alphavirus; WEEV complex.
REFERENCE   1
  AUTHORS   Yuferov,V.P., Volchkov,V.E., Netesov,S.V., Safronov,P.F. and
            Uryvaev,L.V.
  TITLE     Nucleotide sequence of the 26S mRNA and genes of nonstructureal
            proteins NSP2 and NSP3 of western Encephalitis virus strain 5614
  JOURNAL   Virusologia 28, 86-98 (1992)
REFERENCE   2  (bases 1 to 4498)
  AUTHORS   Uryvaev,L.V., Volchkov,V.E., Iuferov,V.P., Samokhvalov,E.I.,
            Lebedev,A.I., Safronov,P.F. and Netesov,S.V.
  TITLE     Primary structure of proteins of the nsP2 and nsP3 polymerase
            complex confirm the recombinant nature of western encephalitis
            virus
  JOURNAL   Dokl. Akad. Nauk SSSR 335, 813-818 (1994)
REFERENCE   3  (bases 1 to 4498)
  AUTHORS   Volchkov,V.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-1993) V.E. Volchkov, All-Union Institute of
            Molecular, Biology of NPO Vector, Koltsovo, Novosibirsk region,
            633159, RUSSIA
FEATURES             Location/Qualifiers
     source          1. .4498
                     /organism="Western equine encephalomyelitis virus"
                     /mol_type="genomic RNA"
                     /strain="5614"
                     /db_xref="taxon:11039"
     CDS             <1. .>4498
                     /codon_start=2
                     /transl_except=(pos:4253. .4255,aa:OTHER)
                     /product="nonstructural polyprotein"
                     /protein_id="CAA52868.2"
                     /db_xref="GI:28400593"
                     /db_xref="GOA:Q89241"
                     /db_xref="SPTREMBL:Q89241"
                     /translation="TIKKVPAVFDSFVIPRLTSHGPDIGLPRRSRLLLEPTVKPAPAI
                     TMADVEHLRGLQQEAEEVAAAEEIREALPPLLPEIEKETVEAEVDLIMQEAGAGSVET
                     PRGHIRVTSYPGEEKIGSYAILSPQAVLNSEKLKCIHPLAEQVLVMTHKGRAGRYKVE
                     PYHGKVIVPEGTAVPVQDFQALSESATIVFNEREFVNRYLHHIAINGGALNTDEEYYK
                     TVKTQDTDSEYVFDIDARKCVKREDAGPLCLTGDLVDPPFHEFAYESLKTRPAAPHKV
                     PTIGVYGVPGSGKSGIIKSAVTKKDLVVSAKKENCAEIIRDVRRMRRMDVAARTVDSV
                     LLNGVKHPVNTLYIDEAFACHAGTLLALIAIVKPKKVVLCGDPKQCGFFNMMCLKVHF
                     NHDICTEVYHKSISRRCTQTVTAIVSTLFYDKRMKTVNPCADKIIIDTTGTTKPHKDD
                     LILTCFRGWVKQLQIDYKNHEIMTAAASQGLTRKGVYAVRYKVNENPLYSQTSEHVNV
                     LVTRTEKRIVWKTLAGDPWIKTFTAKYPGDFTASLDDWQREHDAIMARVLDKPQTADV
                     FQNKVNVCWAKALEPVLATANIVLTRQQWETLHPFKHDRAYSPEMALNFFCTRFFGVD
                     LDSGLFSAPTVALTYRDQHWDNSPGKNMYGLNREVAKELSRRYPCITKAVDTGRVADI
                     RNNTIKDYSPTINVVPLNRRLPHSLIVDHKGQGTTDHSGFLSKMNGKSVLVIGDPISI
                     PGKKVESMGPLPTNTIRCDLDLGIPSHVGKYDIIFVNVRTPYKNHHYQQCEDHAIHHS
                     MLTCKAVHHLNTGGTCVAIGYGLADRATENIITAVARSFRFTRVCQPKNTAENTEVLF
                     VFFGKDNGNHTHDQDRLGVVLDNIYQGSTRYEAGRAPAYRVIRGDISKSADQAIVNAA
                     NSKGQPGSGVCGALYRKWPAAFDRQPIAVGTARLVKHEPLIIHAVGPNFSKMPEREGD
                     LKLAAAYMSIASIVNAERITKISVPLLSTGIYSGGKDRVMQSLHHLFTAFDTTDADVT
                     IYCLDKQWETRITEAIHRKESVEILDDDDKPVDIDLVRVHPNSSLAGRPGYSVNEGKL
                     YSYLEGTRFHQTAKDIAEIHAMWPNKSEANEQICLYILGESMSSIRSKCPVEESEASA
                     PPHTLPCLCNYAMTAERVYRLRSAKKEQFAVCSSFPLPKYRITGVQKLQCSKPVLFSG
                     VVPPAVHPRKYAEIILETPPSPTTTTVICEPTVPERIPSPVISRAPSAESLLSFGGVS
                     FSSSATRSSTAWSDYDRRFVVTADVHQANTSTWSIPSAPGLDVQLPSDDTDSHWSIPS
                     ALGFEVRTPSVQDLTAECARPRGLAEIMQDFNTAPFQFLSDHRPVPAPRRRPIPSPRS
                     TASAPPVPKPRRTKYQQPPGVARAISEAELDEYIRQHSNXRYEAGAYIFSSETGQGHL
                     QQKSVRQCKLQDPIFERAVHEKYYAPRLDLEREKMLQKKLQLCASEGNRSRYQSGRPI
                     PGLQR"
     mat_peptide     <1. .292
                     /product="NSP1"
     mat_peptide     293. .2674
                     /product="NSP2"
     mat_peptide     2675. .4273
                     /product="NSP3"
     mat_peptide     4274. .>4498
                     /product="NSP4"
BASE COUNT     1279 a   1116 c   1125 g    978 t
ORIGIN
```

```
  Query Match             70.7%;  Score 3536;  DB 14;  Length 4498;
  Best Local Similarity   97.9%;  Pred. No. 0;
  Matches 3594;  Conservative    0;  Mismatches   75;  Indels    3;  Gaps    1;

Qy       1332 AACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCA 1391
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCA 60

Qy       1392 CGGGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACTGTCAAACC 1451
             |||||  |||||  ||  | |  ||||||  || ||||||||||||||||||||||||||
Db         61 CGGGCCTGATATAGGGCTTCCGCGTAGGTCCAGGCTGCTGCTTGAACCAACTGTCAAACC 120

Qy       1452 CGCACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGA 1511
             |||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
```


Db 121 CGCACCGGCTATTACAATGGCCGATGTGGAGCACTTGCGTGGCTTACAGCAAGAAGCTGA 180
Qy 1512 AGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGA 1571
Db 181 AGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGA 240
Qy 1572 AAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGAGCAGGTAGCGT 1631
Db 241 AAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGAGCAGGTAGCGT 300
Qy 1632 GGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAGATTGGGTC 1691
Db 301 GGAGACACCACGAGGACACATCAGGGTAACAAGTTACCCCGGTGAAGAAAAGATTGGGTC 360
Qy 1692 TTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGTATCCACCC 1751
Db 361 TTACGCTATACTTTCACCCCAGGCGGTTTTGAATAGTGAAAAACTGAAGTGTATCCACCC 420
Qy 1752 ATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATACAAAGTCGA 1811
Db 421 ATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCTGGGCGATACAAAGTCGA 480
Qy 1812 GCCATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAAGACTTCCA 1871
Db 481 GCCATACCACGGTAAGGTCATTGTACCAGAAGGGACAGCGGTCCCTGTTCAAGACTTTCA 540
Qy 1872 GGCATTGAGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAACAGATACCT 1931
Db 541 GGCACTGAGTGAGAGCGCCACGATCGTTTTCAACGAGAGGGAGTTCGTAAACAGATATTT 600
Qy 1932 GCACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTATAAGACTGT 1991
Db 601 GCACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTATAAGACTGT 660
Qy 1992 AAAGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAGTGTGTTAA 2051
Db 661 AAAGACTCAGGACACAGACTCAGAATACGTCTTTGACATTGACGCACGAAAGTGTGTTAA 720
Qy 2052 GCGAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCATTTCACGA 2111
Db 721 GCGAGAAGACGCAGGTCCCTTATGCCTAACTGGTGATCTGGTAGATCCACCATTTCACGA 780
Qy 2112 GTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCAACCATCGG 2171
Db 781 GTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCAACCATCGG 840
Qy 2172 AGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTGACTAAGAA 2231
Db 841 AGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTGACTAAGAA 900
Qy 2232 AGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGATGTAAGGAG 2291
Db 901 AGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATTATCAGGGATGTAAGGAG 960
Qy 2292 GATGAGACGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAATGGGGTTAA 2351
Db 961 GATGAGACGTATGGATGTTGCTGCTAGGACTGTTGATTCAGTGCTTCTAAATGGGGTTAA 1020
Qy 2352 GCACCCCGTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGGACGCTGCT 2411
Db 1021 GCACCCCGTTAACACCCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGGACGCTGCT 1080
Qy 2412 GGCACTGATTGCCATCGTCAAACCTAAGAAAGTGGTATTGTGCGGGGACCCAAAACAATG 2471
Db 1081 GGCACTGATTGCCATCGTCAAACCTAAGAAAGTGGTATTGTGCGGGGACCCAAAACAATG 1140
Qy 2472 CGGCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGCACTGAAGT 2531
Db 1141 CGGCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGCACTGAGGT 1200
Qy 2532 GTACCATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATCGTCTCCACGCT 2591
Db 1201 GTACCACAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATTGTCTCCACGCT 1260
Qy 2592 CTTCTACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATCATCATAGATAC 2651
Db 1261 CTTCTACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGACAAAATCATCATAGATAC 1320
Qy 2652 CACAGGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGAGGATGGGT 2711
Db 1321 CACAGGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGAGGATGGGT 1380
Qy 2712 GAAACAGCTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCTGCATCGCAAGG 2771
Db 1381 GAAACAGCTACAGATTGACTATAAAAATCATGAAATCATGACTGCGGCTGCATCGCAAGG 1440
Qy 2772 ACTTACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCACTCTACTC 2831
Db 1441 ACTTACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCACTCTACTC 1500
Qy 2832 GCAGACTTCTGAGCACGTGAACGTGTTACTTACACGCACAGAAAAACGCATTGTCTGGAA 2891
Db 1501 GCAGACTTCTGAGCACGTGAACGTGTTAGTTACACGCACAGAAAAACGCATTGTCTGGAA 1560
Qy 2892 GACGCTAGCTGGTGATCCCTGGATAAAGACACTTACAGCTAAATATCCCGGGGATTTCAC 2951
Db 1561 GACGCTAGCTGGTGACCCCTGGATAAAGACATTTACAGCTAAATACCCCGGGGATTTCAC 1620
Qy 2952 GGCTTCATTGGACGACTGGCAGCGCGAACACGACGCCATTATGGCACGCGTTCTTGATAA 3011
Db 1621 GGCTTCATTGGACGACTGGCAGCGCGAACACGACGCTATTATGGCACGCGTTCTTGATAA 1680
Qy 3012 GCCGCAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAGGCTTTAGA 3071
Db 1681 ACCGCAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAGGCTTTAGA 1740
Qy 3072 GCCAGTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAGACGTTGCACCC 3131
Db 1741 GCCAGTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAGACGTTGCACCC 1800
Qy 3132 ATTCAAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTTTGCACCAGGTT 3191
Db 1801 ATTCAAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTTTGCACCAGGTT 1860
Qy 3192 CTTTGGAGTAGACCTGGACAGTGGGTTATTTTCCGCTCCTACCGTCGCACTTACTTACAG 3251
Db 1861 CTTTGGTGTAGACCTGGACAGTGGGTTGTTTTCCGCTCCTACCGTCGCACTTACTTACAG 1920
Qy 3252 GGATCAGCACTGGGATAACTCGCCAGGGAAGAACATGTATGGGCTTAATAGAGAGGTAGC 3311
Db 1921 GGATCAGCACTGGGACAACTCGCCAGGGAAGAACATGTATGGGCTTAATAGAGAGGTAGC 1980
Qy 3312 AAAGGAGTTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACAGGCAGGGTAGC 3371
Db 1981 AAAGGAGCTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACAGGCAGGGTAGC 2040
Qy 3372 TGATATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTGGTTCCATTAAA 3431
Db 2041 TGATATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTGGTTCCATTAAA 2100
Qy 3432 TCGCCGGTTGCCCCACTCGTTGATCGTTGACCACAAAGGACAGGGTACAACTGATCACAG 3491
Db 2101 TCGCCGGTTGCCCCACTCGCTGATCGTTGACCACAAAGGACAGGGTACAACTGATCACAG 2160
Qy 3492 CGGATTCCTATCTAAGATGAAGGGCAAATCTGTGTTGGTGATCGGCGATCCTATCAGCAT 3551
Db 2161 TGGATTCCTATCTAAGATGAATGGCAAATCTGTGTTGGTGATCGGCGATCCTATCAGCAT 2220
Qy 3552 TCCAGGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATACCATCAGGTGTGATCT 3611
Db 2221 TCCTGGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATACCATCAGGTGTGATCT 2280
Qy 3612 CGATTTGGGAATACCTAGCCATGTCGGTAAATATGACATTATCTTTGTCAATGTTAGGAC 3671
Db 2281 CGATTTGGGAATACCTAGCCATGTCGGTAAATATGACATTATATTTGTCAATGTTAGGAC 2340

```
Qy       3672 CCCGTACAGGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCACCACAGCATGCT 3731
             ||| |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CCCATACAAGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCACCACAGCATGCT 2400

Qy       3732 AACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCTATAGGGTATGG 3791
             |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       2401 AACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCCATAGGGTATGG 2460

Qy       3792 GCTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCACGCTCATTTAGGTTTAC 3851
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db       2461 ACTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCTCGCTCATTTAGGTTTAC 2520

Qy       3852 CCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTCGTGTTCTTCGG 3911
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 CCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTCGTGTTCTTCGG 2580

Qy       3912 CAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTGCTTGACAACAT 3971
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 CAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTGCTTGACAACAT 2640

Qy       3972 CTATCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGTACAGAGTGATCAGAGG 4031
             ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 CTACCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGTACAGAGTGATCAGAGG 2700

Qy       4032 TGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGCAAAGGTCAACC 4091
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 TGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGCAAAGGTCAACC 2760

Qy       4092 AGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTTGATAGACAGCC 4151
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 AGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTTGATAGACAGCC 2820

Qy       4152 AATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACCGCTCATCATACATGCTGTAGG 4211
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 AATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACCGCTCATCATACATGCTGTAGG 2880

Qy       4212 ACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTAAGCTCGCAGCTGCCTACAT 4271
             ||| |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       2881 ACCGAATTTTTCTAAGATGCCGGAACGAGAGGGCGACCTTAAGCTCGCAGCTGCCTACAT 2940

Qy       4272 GAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTACCGCTACTGTC 4331
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       2941 GAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACTAAAATATCAGTACCGCTACTGTC 3000

Qy       4332 AACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCATCACCTGTTCAC 4391
             ||| ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       3001 AACTGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCATCACTTGTTCAC 3060

Qy       4392 TGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAACAATGGGAGAC 4451
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 TGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAACAATGGGAGAC 3120

Qy       4452 CAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATTCTG---GATGATGACAA 4508
             |||||||| |||||||||||||||||||||||||||||||||| |||   ||||||||||
Db       3121 CAGGATAACCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATACTGGATGATGATGACAA 3180

Qy       4509 GCCAGTAGACATTGACTTGGTCAGGGTCCACCCAAACAGCTCTTTGGCAGGCAGACCAGG 4568
             |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       3181 GCCAGTAGACATCGACTTGGTCAGGGTCCACCCAAACAGCTCTTTGGCAGGCAGACCAGG 3240

Qy       4569 TTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTGGAAGGTACACGATTCCATCAGAC 4628
             |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       3241 TTACTCCGTCAATGAGGGCAAGCTGTATTCATACCTGGAAGGTACACGATTCCATCAGAC 3300

Qy       4629 CGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCCAACAAATCTGAGGCTAATGAGCA 4688
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 CGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCCAACAAATCTGAGGCTAATGAGCA 3360

Qy       4689 GATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGCCCAGTAGA 4748
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 GATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGCCCAGTAGA 3420

Qy       4749 GGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTACGCTATGAC 4808
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 GGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTACGCTATGAC 3480

Qy       4809 GGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTATGCTCATC 4868
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 GGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTATGCTCATC 3540

Qy       4869 ATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGCAAACCAGT 4928
             ||||| |||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db       3541 ATTCCCGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAATGCAGCAAACCAGT 3600

Qy       4929 CCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGCGGAAATAATTCT 4988
             |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 CCTGTTCTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGCGGAAATAATTCT 3660

Qy       4989 AGAAACGCCACC 5000
             ||||||||||||
Db       3661 AGAAACGCCACC 3672
```

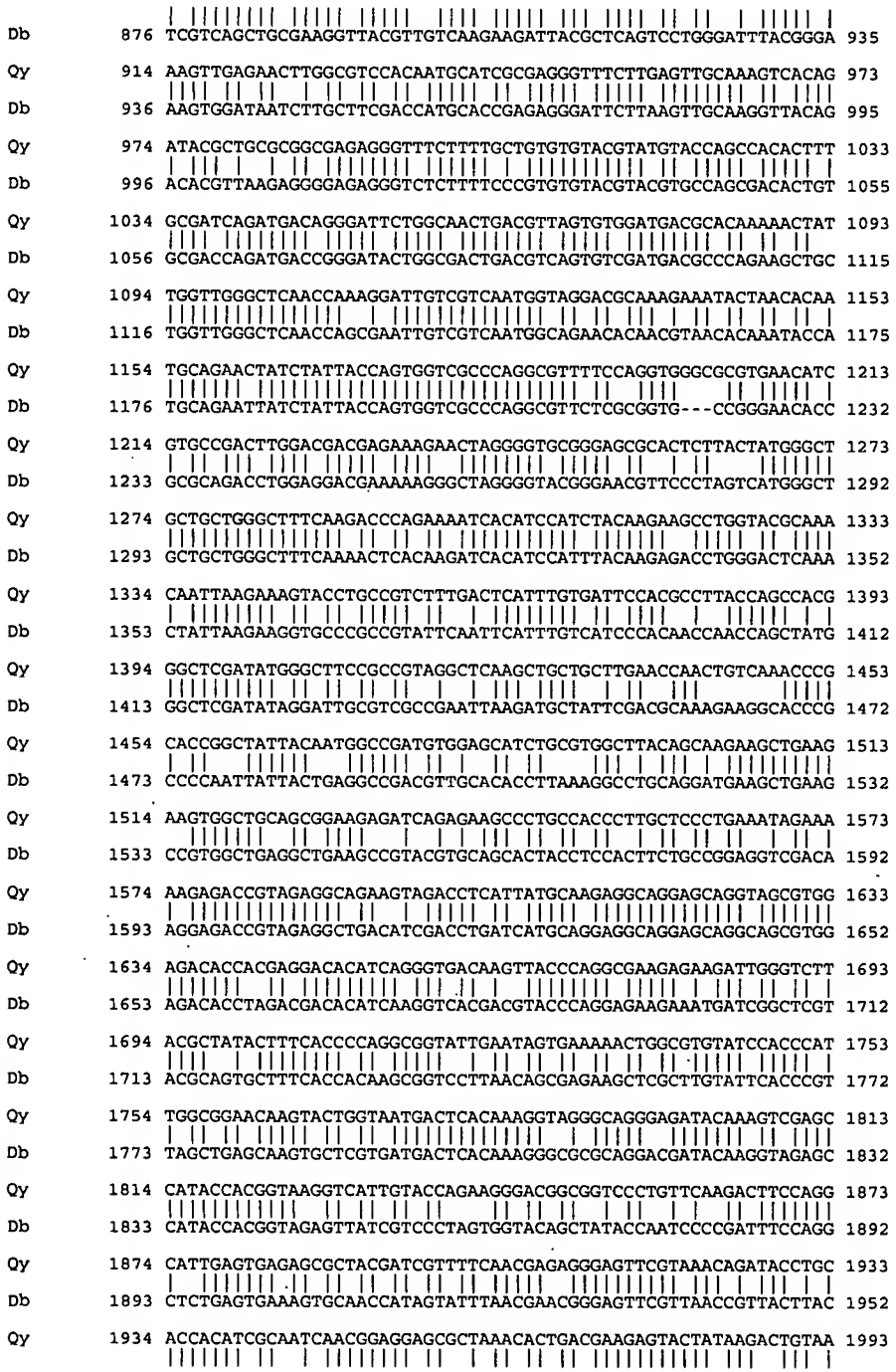
RESULT 4
U01034

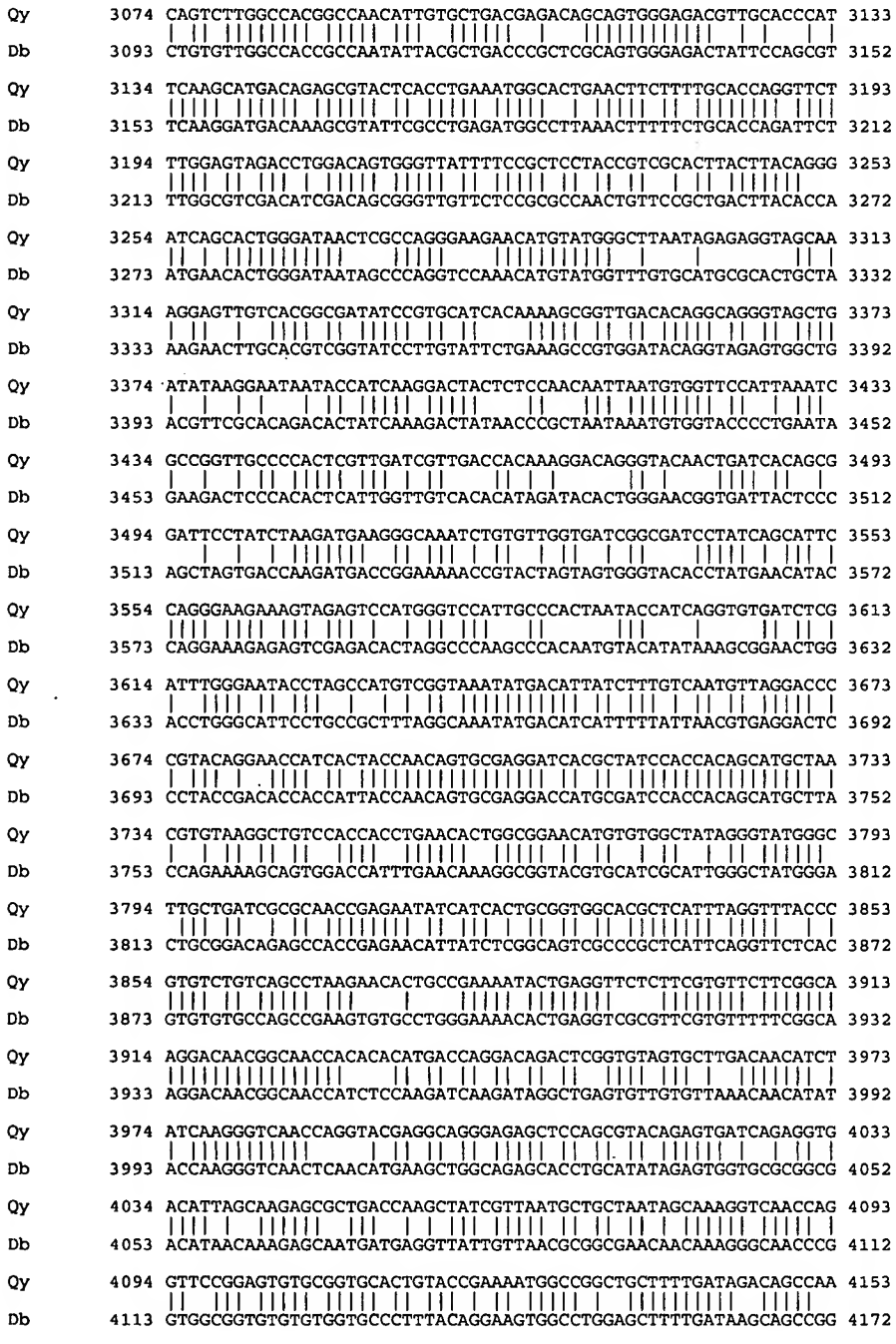
```
LOCUS       U01034                 11678 bp    RNA     linear   VRL 02-MAR-2000
DEFINITION  Eastern equine encephalomyelitis virus North American antigenic
            variety nonstructural polyprotein and structural polyprotein genes,
            complete cds.
ACCESSION   U01034
VERSION     U01034.1  GI:393006
KEYWORDS    nsP1; nsP2; nsP3; nsP4; capsid; E3; E2; 6K; E1.
SOURCE      Eastern equine encephalitis virus
  ORGANISM  Eastern equine encephalitis virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
            Alphavirus; EEEV complex.
REFERENCE   1  (bases 1 to 11678)
  AUTHORS   Weaver,S.C., Hagenbaugh,A., Bellew,L., Netesov,S.V., Volchkov,V.E.,
            Chang,G.J., Clarke,D.K., Gousset,L., Scott,T.W., Trent,D.W. and
            Holland,J.J.
  TITLE     A comparison of the nucleotide sequences of eastern and western
            equine encephalomyelitis viruses with those of other alphaviruses
            and related RNA viruses
  JOURNAL   Virology 197 (1), 375-390 (1993)
  MEDLINE   94025587
   PUBMED   8105605
REFERENCE   2  (bases 1 to 11678)
  AUTHORS   Weaver,S.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-1993) Scott C. Weaver, Biology, University of
            California, San Diego, La Jolla, CA, 92037-0116
FEATURES             Location/Qualifiers
     source          1. .11678
                     /organism="Eastern equine encephalitis virus"
                     /mol_type="genomic RNA"
                     /strain="North American antigenic variety"
                     /isolate="82V2137"
                     /db_xref="taxon:11021"
                     /clone="pEE14 pEE72 pEE59 pEE700 pEE1900 pEE700"
     5'UTR           1. .46
     CDS             47. .7528
                     /note="termination codon at position 5681. .5683"
                     /codon_start=1
                     /transl_except=(pos:5681. .5683,aa:OTHER)
                     /product="nonstructural polyprotein"
                     /protein_id="AAC53734.1"
                     /db_xref="GI:393007"
                     /translation="MEKVHVDLDADSPFVKSLQRCFPHFEIEATQVTDNDHANARAFS
                     HLATKLIEGEVDTDQVILDIGSAPVRHTHSKHKYHCICPMKSAEDPDRLYRYADKLRK
                     SDVTDKCIASKAADLLTVMSTPDAETPSLCMHTDSTCRYHGSVAVYQDVYAVHAPTSI
                     YYQALKGVRTIYWIGFDTTPFMYKNMAGAYPTYNTNWADESVLEARNIGLGSSDLHEK
                     SFGKVSIMRKKKLQPTNKVIFSVGSTIYTEERILLRSWHLPNVFHLKGKTSFTGRCNT
                     IVSCEGYVVKKITLSPGIYGKVDNLASTMHREGFLSCKVTDTLRGERVSFPVCTYVPA
                     TLCDQMTGILATDVSVDDAQKLLVGLNQRIVVNGRTQRNTNTMQNYLLPVVAQAFSRC
                     REHRADLEDEKGLGVRERSLVMGCCWAFKTHKITSIYKRPGTQTIKKVPAVFNSFVIP
                     QPTSYGLDIGLRRRIKMLFDAKKAPAPIITEADVAHLKGLQDEAEAVAEAEAVRAALP
```

```
                PLLPEVDKETVEADIDLIMQEAGAGSVETPRRHIKVTTYPGEEMIGSYAVLSPQAVLN
                SEKLACIHPLAEQVLVMTHKGRAGRYKVEPYHGRVIVPSGTAIPIPDFQALSESATIV
                FNEREFVNRYLHHIAVNGGALNTDEEYYKVVKSTETDSEYVFDIDAKKCVKKGDAGPM
                CLVGELVDPPFHEFAYESLKTRPAAPHKVPTIGVYGVPGSGKSGIIKSAVTKRDLVVS
                AKKENCMEIIKDVKRMRGMDIAARTVDSVLLNGVKHSVDTLYIDEAFACHAGTLLALI
                AIVKPKKVVLCGDPKQCGFFNLMCLKVHFNHEICTEVYHKSISRRCTKTVTSIVSTLF
                YDKRMRTVNPCNDKIIIDTTSTTKPLKDDIILTCFRGWVKQLQIDYKNHEIMTAAASQ
                GLTRKGVYAVRYKVNENPLYAQTSEHVNVLLTRTEKRIVWKTLAGDPWIKTLTASYPG
                NFTATLEEWQAEHDAIMAKILETPASSDVFQNKVNVCWAKALEPVLATANITLTRSQW
                ETIPAFKDDKAYSPEMALNFFCTRFFGVDIDSGLFSAPTVPLTYTNEHWDNSPGPNMY
                GLCMRTAKELARRYPCILKAVDTGRVADVRTDTIKDYNPLINVVPLNRRLPHSLVVTH
                RYTGNGDYSQLVTKMTGKTVLVVGTPMNIPGKRVETLGPSPQCTYKAELDLGIPAALG
                KYDIIFINVRTPYRHHHYQQCEDHAIHHSMLTRKAVDHLNKGGTCIALGYGTADRATE
                NIISAVARSFRFSRVCQPKCAWENTEVAFVFFGKDNGNHLQDQDRLSVVLNNIYQGST
                QHEAGRAPAYRVVRGDITKSNDEVIVNAANNKGQPGGGVCGALYRKWPGAFDKQPVAT
                GKAHLVKHSPNVIHAVGPNFSRLSENEGDQKLSEVYMDIARIINNERFTKVSIPLLST
                GIYAGGKDRVMQSLNHLFTAMDTTDADITIYCLDKQWESRIKEAITRKESVEELTEDD
                RPVDIELVRVHPLSSLAGRPGYSTTEGKVYSYLEGTRFHQTAKDIAEIYAMWPNKQEA
                NEQICLYVLGESMNSIRSKCPVEESEASSPPHTIPCLCNYAMTAERVYRLRMAKNEQF
                AVCSSFQLPKYRITGVQKIQCSKPVIFSGTVPPAIHPRKFASVTVEDTPVVQPERLVP
                RRPAPPVPVPARIPSPPCTSTNGSTTSIQSLGEDQSASASSGAEISVDHVSLWSIPSA
                TGFDVRTSSSLSLEQPTFPTMVVEAEIHASQGSLWSIPSITGSETRVPSPPSQDSRPS
                TPSASGSHTSVDLITFDSVAEILEDFSRSPFQFLSEIKPIPAPRTRVNNMSRSADTIK
                PIPKPRKCQVKYTQPPGVARAISAAEFDEFVRRHSNXRYEAGAYIFSSETGQGHLQQK
                STRQCKLQYPILERSVHEKFYAPRLDLEREKLLQKKLQLCASEGNRSRYQSRKVENMK
                AITVERLLQGIGSYLSAEPQPVECYKVTYPAPMYSSTASNSFSSAEVAVKVCNLVLQE
                NFPTVASYNITDEYDAYLDMVDGASCCLDTATFCPAKLRSFPKKHSYLRPEIRSAVPS
                PIQNTLQNVLAAATKRNCNVTQMRELPVLDSAAFNVECFKKYACNDEYWDFYKTNPIR
                LTTENVTQYVTKLKGPKAAALFAKTHNLQPLHEIPMDRFVMDLKRDVKVTPGTKHTEE
                RPKVQVIQAADPLATAYLCGIHRELVRRLNAVLLPNIHTLFDMSAEDFDAIIAEHFQF
                GDAVLETDIASFDKSEDDAIAMSALMILEDLGVDQALLNLIEAAFGNITSVHLPTGTR
                FKFGAMMKSGMFLTLFINTVVNIMIASRVLRERLTTSPCAAFIGDDNIVKGVTSDALM
                AERCATWLNMEVKIIDAVVGVKAPYFCGGFIVVDQITGTACRVADPLKRLFKLGKPLP
                LDDDQDVDRRRALHDEAARWNRIGITEELVKAVESRYEVNYVSLIITALTTLASSVSN
                FKHIRGHPITLYG"
mat_peptide     47. .1642
                /product="nsP1"
mat_peptide     1643. .4024
                /product="nsP2"
mat_peptide     4025. .5701
                /product="nsP3"
mat_peptide     5702. .7525
                /product="nsP4"
mRNA            7534. .11678
CDS             7592. .11320
                /codon_start=1
                /product="structural polyprotein"
                /protein_id="AAC53735.1"
                /db_xref="GI:393008"
                /translation="MFPYPTLNYPPMAPINPMAYRDPNPPRRRWRPFRPPLAAQIEDL
                RRSIANLTLKQRAPNPPAGPPAKRKKPAPKPKPAQTKKKRPPPPAKKQKRKPKPGKRQ
                RMCMKLESDKTFPIMLNGQVNGYACVVGGRVFKPLHVEGRIDNEQLAAIKLKKASIYD
                LEYGDVPQCMKSDTLQYTSDKPPGFYNWHHGAVQYENNRFTVPRGVGGKGDSGRPILD
                NKGRVVAIVLGGVNEGSRTALSVVTWNQKGVTVKDTPEGSEPWSLATVMCVLANITFP
                CDQPPCMPCCYEKNPHETLTMLEQNYDSRAYDQLLDAAVKCNARRTRRDLDTHFTQYK
                LARPYIADCPNCGHSRCDSPIAIEEVRGDAHAGVIRIQTSAMFGLKTDGVDLAYMSFM
                NGKTQKSIKIDNLHVRTSAPCSLVSHHGYYILAQCPPGDTVTVGFHDGPNRHTCTVAH
                KVEFRPVGREKYRHPPEHGVELPCNRYTHKRADQGHYVEMHQPGLVADHSLLSIHSAK
                VKITVPSGAQVKYYCKCPDVREGITSSDHTTTCTDVKQCRAYLIDNKKWVYNSGRLPR
                GEGDTFKGKLHVPFVPVKAKCIATLAPEPLVEHKHRTLILHLHPDHPTLLTTRSLGSD
                ANPTRQWIERPTTVNFTVTGEGLEYTWGNHPPKRVWAQESGEGNPHGWPHEVVVYYYN
                RYPLTTIIGLCTCVAIIMVSCVTSVWLLFRTRNLCITPYKLAPNAQVPILLALLCCIK
                PTRADDTLQVLNYLWNNNQNFFWMQTLIPLAALIVCMRMLRCLFCCGPAFLLVCGALG
                AAAYEHTAVMPNKVGIPYKALVERPGYAPVHLQIQLVNTRIIPSTNLEYITCKYKTKV
                PSPVVKCCGATQCTSKPHPDYQCQVFTGVYPFMWGGAYCFCDTENTQMSEAYVERSEE
                CSIDHAKAYKVHTGTVQAMVNITYGSVSWRSADVYVNGETPAKIGDAKLIIGPLSSAW
                SPFDNKVVVYGHEVYNYDFPEYGTGKAGSFGDLQSRTSTSNDLYANTNLKLQRPQAGI
                VHTPFTQAPSGFERWKRDKGAPLNDVAPFGCSIALEPLRAENCAVGSIPISIDIPDAA
                FTRISETPTVSDLECKITECTYASDFGGIATVAYKSSKAGNCPIHSPSGVAVIKENDV
                TLAESGSFTFHFSTANIHPAFKLQVCTSAVTCKGDCKPPKDHIVDYPAQHTESFTSAI
                SATAWSWLKVLVGGTSAFIVLGLIATAVVALVLFFHRH"
mat_peptide     7592. .8374
                /product="capsid"
mat_peptide     8375. .8563
                /product="E3"
mat_peptide     8564. .9823
                /product="E2"
mat_peptide     9824. .9994
                /product="6K"
mat_peptide     9995. .11317
                /product="E1"
3'UTR           11321. .11678
BASE COUNT      3314 a  2894 c  2809 g  2661 t
ORIGIN
```

Query Match 55.3%; Score 2766; DB 14; Length 11678;
Best Local Similarity 72.6%; Pred. No. 0;
Matches 3605; Conservative 0; Mismatches 1355; Indels 6; Gaps 2;

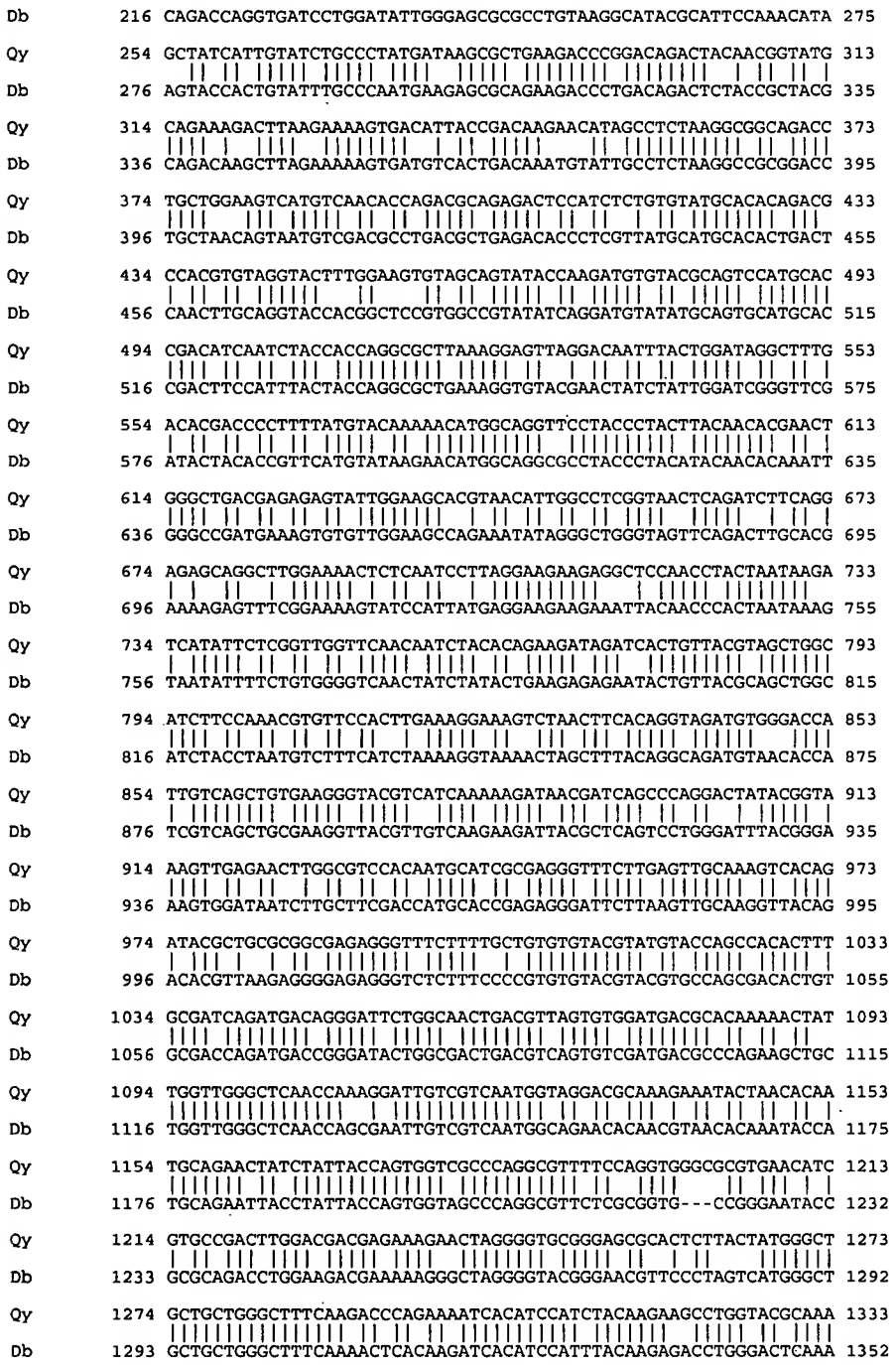
```
Qy      14 ATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGT 73
           | | |||||| ||||| | | |||| |||||||||||| || |||||||| |  |||||||
Db      36 ACCTATCCAAAATGGAGAAAGTTCATGTTGACTTAGACGCAGACAGCCCATTCGTCAAGT 95

Qy      74 CGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACC 133
           | | ||  |    |||||||| |||||||| ||||||| ||||||||||||||||||||
Db      96 CACTGCAAAGATGCTTTCCACATTTGAGATAGAAGCAACGCAGGTCACTGACAATGACC 155

Qy     134 ATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACC 193
           |||| ||||| || |||||||||||  | || || |||||||||||| | |||||  ||
Db     156 ATGCTAATGCTAGGGCGTTTTCGCACCTAGCTACTAAGCTCATTGAGGGAGAAGTGGATA 215

Qy     194 GGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACC 253
           |||||  || ||| |||| ||||| || ||||| || || ||  | |||||||  ||
Db     216 CAGACCAGGTGATCCTGGATATTGGGAGCGCGCCTGTAAGGCACACGCATTCCAAACATA 275

Qy     254 GCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGGTATG 313
           || || ||||| ||||| |||| ||||| |||||||| ||||||||  | || || |
Db     276 AGTACCACTGTATTTGCCCAATGAAGAGCGCAGAAGACCCTGACAGACTCTACCGCTACG 335

Qy     314 CAGAAAGACTTAAGAAAAGTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACC 373
           |||| |  ||||  || |||||  | || ||||     || ||||||||||| || ||||
Db     336 CAGACAAGCTTAGAAAGAGTGATGTCACTGACAAATGTATTGCCTCTAAGGCCGCGGACC 395

Qy     374 TGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACG 433
           ||||   ||| ||||| || || ||||| ||||| || ||  | || |||||||| |||
Db     396 TGCTAACAGTAATGTCGACGCCTGACGCTGAGACACCCTCGTTATGCATGCACACTGACT 455

Qy     434 CCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCAC 493
           | || || ||||||   ||    || || ||||| || ||||| || ||||| |||||||
Db     456 CAACTTGCAGGTACCACGGCTCCGTGGCCGTATATCAGGATGTATATGCAGTGCATGCAC 515

Qy     494 CGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTG 553
           |||| || || ||| |||||||||| ||||| ||  | || || || ||||| || || |
Db     516 CGACTTCCATTTACTACCAGGCGCTGAAAGGTGTACGAACTATCTATTGGATCGGGTTCG 575

Qy     554 ACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACT 613
           | || || || || ||||| || ||||||||||  |||||||||| |||||||| || |
Db     576 ATACTACACCGTTCATGTATAAGAACATGGCAGGCGCCTACCCTACATACAACACAAATT 635

Qy     614 GGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGG 673
           |||| || || || || ||||||||  | || || || || ||||  |||||  | || |
Db     636 GGGCCGATGAAAGTGTGTTGGAAGCCAGAAATATAGGGCTGGGTAGTTCAGACTTGCACG 695

Qy     674 AGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGA 733
           | |  ||  | |||||| | || ||  | |||||||||||    | ||||| || |||||
Db     696 AAAAGAGTTTCGGAAAAGTATCCATTATGAGGAAGAAGAAATTACAACCCACCAATAAAG 755

Qy     734 TCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGC 793
           | ||||| || || || ||||| || || || ||||| ||| |||||||||| || ||||
Db     756 TAATATTTTCTGTGGGGTCAACTATTTATACTGAAGAGAGAATACTGTTACGCAGTTGGC 815

Qy     794 ATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCA 853
           |||| || || || || ||  | ||||| ||  ||| ||| ||||| ||||||   ||||
Db     816 ATCTACCTAATGTTTTTCATCTAAAAGGTAAAACTAGCTTTACAGGCAGATGTAACACCA 875

Qy     854 TTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTA 913
```

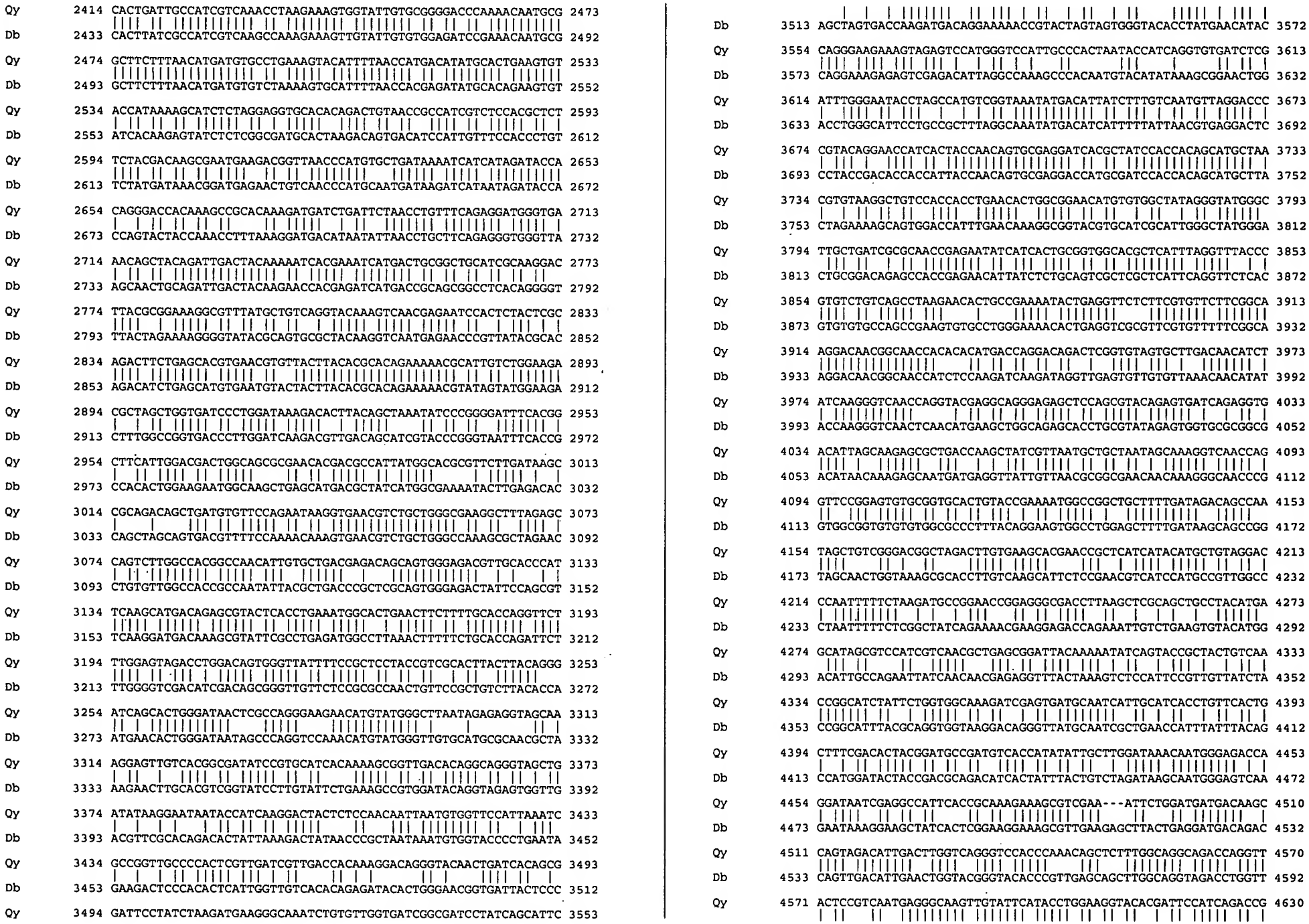

Db 876 TCGTCAGCTGCGAAGGTTACGTTGTCAAGAAGATTACGCTCAGTCCTGGGATTTACGGGA 935
Qy 914 AAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAG 973
Db 936 AAGTGGATAATCTTGCTTCGACCATGCACCGAGAGGGATTCTTAAGTTGCAAGGTTACAG 995
Qy 974 ATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTT 1033
Db 996 ACACGTTAAGAGGGGAGAGGGTCTCTTTTCCCGTGTGTACGTACGTGCCAGCGACACTGT 1055
Qy 1034 GCGATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTAT 1093
Db 1056 GCGACCAGATGACCGGGATACTGGCGACTGACGTCAGTGTCGATGACGCCCAGAAGCTGC 1115
Qy 1094 TGGTTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAA 1153
Db 1116 TGGTTGGGCTCAACCAGCGAATTGTCGTCAATGGCAGAACACAACGTAACACAAATACCA 1175
Qy 1154 TGCAGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATC 1213
Db 1176 TGCAGAATTATCTATTACCAGTGGTCGCCCAGGCGTTCTCGCGGTG---CCGGGAACACC 1232
Qy 1214 GTGCCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCT 1273
Db 1233 GCGCAGACCTGGAGGACGAAAAAGGGCTAGGGGTACGGGAACGTTCCCTAGTCATGGGCT 1292
Qy 1274 GCTGCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCTGGTACGCAAA 1333
Db 1293 GCTGCTGGGCTTTCAAAACTCACAAGATCACATCCATTTACAAGAGACCTGGGACTCAAA 1352
Qy 1334 CAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCACG 1393
Db 1353 CTATTAAGAAGGTGCCCGCCGTATTCAATTCATTTGTCATCCCACAACCAACCAGCTATG 1412
Qy 1394 GGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACTGTCAAACCCG 1453
Db 1413 GGCTCGATATAGGATTGCGTCGCCGAATTAAGATGCTATTCGACGCAAAGAAGGCACCCG 1472
Qy 1454 CACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAG 1513
Db 1473 CCCCAATTATTACTGAGGCCGACGTTGCACACCTTAAAGGCCTGCAGGATGAAGCTGAAG 1532
Qy 1514 AAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAA 1573
Db 1533 CCGTGGCTGAGGCTGAAGCCGTACGTGCAGCACTACCTCCACTTCTGCCGGAGGTCGACA 1592
Qy 1574 AAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGAGCAGGTAGCGTGG 1633
Db 1593 AGGAGACCGTAGAGGCTGACATCGACCTGATCATGCAGGAGGCAGGAGCAGGCAGCGTGG 1652
Qy 1634 AGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAGATTGGGTCTT 1693
Db 1653 AGACACCTAGACGACACATCAAGGTCACGACGTACCCAGGAGAAGAAATGATCGGCTCGT 1712
Qy 1694 ACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGTATCCACCCAT 1753
Db 1713 ACGCAGTGCTTTCACCACAAGCGGTCCTTAACAGCGAGAAGCTCGCTTGTATTCACCCGT 1772
Qy 1754 TGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATACAAAGTCGAGC 1813
Db 1773 TAGCTGAGCAAGTGCTCGTGATGACTCACAAAGGGCGCGCAGGACGATACAAGGTAGAGC 1832
Qy 1814 CATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAAGACTTCCAGG 1873
Db 1833 CATACCACGGTAGAGTTATCGTCCCTAGTGGTACAGCTATACCAATCCCCGATTTCCAGG 1892
Qy 1874 CATTGAGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAACAGATACCTGC 1933
Db 1893 CTCTGAGTGAAAGTGCAACCATAGTATTTAACGAACGGGAGTTCGTTAACCGTTACTTAC 1952
Qy 1934 ACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTATAAGACTGTAA 1993
Db 1953 ACCACATTGCCGTTAACGGAGGGGCATTGAATACAGATGAAGAGTACTACAAGGTTGTGA 2012
Qy 1994 AGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAGTGTGTTAAGC 2053
Db 2013 AAAGCACTGAGACAGACTCTGAGTACGTATTTGACATCGACGCAAAGAAGTGCGTGAAGA 2072
Qy 2054 GAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCATTTCACGAGT 2113
Db 2073 AAGGGGATGCCGGACCAATGTGCCTGGTCGGCGAACTAGTTGACCCGCCATTCCACGAAT 2132
Qy 2114 TTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCAACCATCGGAG 2173
Db 2133 TTGCGTACGAGAGTTTAAAAACACGTCCTGCTGCACCACACAAAGTGCCTACCATCGGAG 2192
Qy 2174 TCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTGACTAAGAAAG 2233
Db 2193 TTTATGGAGTCCCAGGTTCCGGAAAGTCTGGTATAATCAAAAGCGCTGTTACCAAGCGTG 2252
Qy 2234 ATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGATGTAAGGAGGA 2293
Db 2253 ATCTGGTAGTCAGTGCAAAGAAAGAAAATTGCATGGAAATCATTAAAGACGTCAAACGTA 2312
Qy 2294 TGAGACGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAATGGGGTTAAGC 2353
Db 2313 TGCGCGGCATGGACATCGCCGCCCGCACAGTGGATTCGGTACTGCTAAATGGGGTAAAAC 2372
Qy 2354 ACCCCGTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGGACGCTGCTGG 2413
Db 2373 ACTCCGTCGACACACTGTACATAGACGAGGCATTCGCTTGCCATGCAGGGACCCTGCTAG 2432
Qy 2414 CACTGATTGCCATCGTCAAACCTAAGAAAGTGGTATTGTGCGGGGACCCAAAACAATGCG 2473
Db 2433 CACTTATTGCCATCGTCAAGCCAAAGAAAGTTGTATTGTGTGGAGATCCGAAACAATGCG 2492
Qy 2474 GCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGCACTGAAGTGT 2533
Db 2493 GCTTCTTTAACTTGATGTGTCTAAAAGTGCATTTTAACCACGAGATATGCACAGAAGTGT 2552
Qy 2534 ACCATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATCGTCTCCACGCTCT 2593
Db 2553 ATCACAAGAGTATTTCTCGGCGATGCACTAAGACAGTGACATCCATCGTTTCCACCCTGT 2612
Qy 2594 TCTACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATCATCATAGATACCA 2653
Db 2613 TCTATGATAAACGGATGAGAACTGTCAACCCATGCAATGATAAGATCATAATAGATACCA 2672
Qy 2654 CAGGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGAGGATGGGTGA 2713
Db 2673 CCAGTACTACCAAACCTTTAAAGGATGACATAATATTAACCTGCTTTAGAGGGTGGGTTA 2732
Qy 2714 AACAGCTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCTGCATCGCAAGGAC 2773
Db 2733 AACAACTGCAGATTGACTACAAGAACCACGAGATCATGACCGCAGCGGCCTCACAGGGGC 2792
Qy 2774 TTACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCACTCTACTCGC 2833
Db 2793 TTACTAGAAAAGGGGTATACGCAGTGCGCTACAAGGTCAATGAGAACCCACTATACGCAC 2852
Qy 2834 AGACTTCTGAGCACGTGAACGTGTTACTTACACGCACAGAAAAACGCATTGTCTGGAAGA 2893
Db 2853 AGACATCAGAGCATGTGAATGTATTACTTACACGCACTGAAAAACGTATAGTATGGAAGA 2912
Qy 2894 CGCTAGCTGGTGATCCCTGGATAAAGACACTTACAGCTAAATATCCCGGGGATTTCACGG 2953
Db 2913 CTTTGGCCGGTGACCCTTGGATCAAGACGTTGACAGCATCGTATCCAGGTAATTTCACCG 2972
Qy 2954 CTTCATTGGACGACTGGCAGCGCGAACACGACGCCATTATGGCACGCGTTCTTGATAAGC 3013
Db 2973 CCACACTGGAAGAATGGCAAGCTGAGCATGACGCTATCATGGCGAAAATACTTGAGACAC 3032
Qy 3014 CGCAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAGGCTTTAGAGC 3073
Db 3033 CAGCTAGCAGCGACGTTTTCCAAAATAAAGTGAACGTCTGCTGGGCCAAAGCGCTAGAAC 3092



RESULT 5
EEEVIRNA
LOCUS EEEVIRNA 11675 bp RNA linear VRL 08-JUL-1992
DEFINITION Eastern Equine Encephalomyelitis Virus RNA sequence.
ACCESSION X63135 X67111
VERSION X63135.1 GI:59185
KEYWORDS 6K protein; C protein; E1; E2; E3; NS1; NS2; NS3; NS4.
SOURCE Eastern equine encephalitis virus
ORGANISM Eastern equine encephalitis virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
Alphavirus; EEEV complex.
REFERENCE 1 (bases 1 to 11675)
AUTHORS Volchkov,V.E., Volchkova,V.A. and Netesov,S.V.
TITLE Complete nucleotide sequence of the Eastern equine
encephalomyelitis virus genome
JOURNAL Mol. Gen. Mikrobiol. Virusol. 5, 8-15 (1991)

```
  MEDLINE   91375524
   PUBMED   1896061
REFERENCE   2  (bases 1 to 11675)
  AUTHORS   Volchkov,V.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-1991) V.E. Volchkov, All Union Institute of
            Molecular Biology, Koltsovo, Novosibirsk Region, 633159, USSR
FEATURES             Location/Qualifiers
     source          1. .11675
                     /organism="Eastern equine encephalitis virus"
                     /mol_type="genomic RNA"
                     /strain="ssp. North American variant"
                     /db_xref="taxon:11021"
                     /cell_line="virus grown on vero cells"
                     /clone_lib="cDNA plasmid clones grown in E.coli JC-5183"
     CDS             47. .5683
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAA44846.1"
                     /db_xref="GI:59187"
                     /db_xref="SPTREMBL:Q66580"
                     /translation="MEKVHVDLDADSPFVKSLQRCFPHFEIEATQVTDNDHANARAFS
                     HLATKLIEGEVDTDQVILDIGSAPVRHTHSKHKYHCICPMKSAEDPDRLYRYADKLRK
                     SDVTDKCIASKAADLLTVMSTPDAETPSLCMHTDSTCRYHGSVAVYQDVYAVHAPTSI
                     YYQALKGVRTIYWIGFDTTPFMYKNMAGAYPTYNTNWADESVLEARNIGLGSSDLHEK
                     SFGKVSIMRKKKLQPTNKVIFSVGSTIYTEERILLRSWHLPNVFHLKGKTSFTGRCNT
                     IVSCEGYVVKKITLSPGIYGKVDNLASTMHREGFLSCKVTDTLRGERVSFPVCTYVPA
                     TLCDQMTGILATDVSVDDAQKLLVGLNQRIVVNGRTQRNTNTMQNYLLPVVAQAFSRC
                     REYRADLEDEKGLGVRERSLVMGCCWAFKTHKITSIYKRPGTQTIKKVPAVFNSFVIP
                     QPTSYGPDIGLRRRIKMLFDAKKAPAPIITEADVAHLKGLQDEAEAVAEAEAVRAALP
                     PLLPEVDKETVEADIDLIMQEAGAGSVETPRRHIKVTTYPGEEMIGSYAVLSPQAVLN
                     SEKLACIHPLAEQVLVMTHKGRAGRYKVEPYHDRVIVPSGTAIPIPDFQALSESATIV
                     FNEREFVNRYLHHIAVNGGALNTDEEYYKVVKSTETDSEYVFDIDAKKCVKKGDAGPM
                     CLVGELVDPPFHEFAYESLKTRPAAPHKVPTIGVYGVPGSGKSGIIKSAVTKRDLVVS
                     AKKENCMEIIKDVKRMRGMDIAARTVDSVLLNGVKHSVDTLYIDEAFACHAGTLLALI
                     AIVKPKKVVLCGDPKQCGFFNMMCLKVHFNHEICTEVYHKSISRRCTKTVTSIVSTLF
                     YDKRMRTVNPCNDKIIIDTTSTTKPLKDDIILTCFRGWVKQLQIDYKNHEIMTAAASQ
                     GFTRKGVYAVRYKVNENPLYAQTSEHVNVLLTRTEKRIVWKTLAGDPWIKTLTASYPG
                     NFTATLEEWQAEHDAIMAKILETPASSDVFQNKVNVCWAKALEPVLATANITLTRSQW
                     ETIPAFKDDKAYSPEMALNFFCTRFFGVDIDSGLFSAPTVPLSYTNEHWDNSPGPNMY
                     GLCMRNAKELARRYPCILKAVDTGRVVDVRTDTIKDYNPLINVVPLNRRLPHSLVVTQ
                     RYTGNGDYSQLVTKMTGKTVLVVGTPMNIPGKRVETLGQSPQCTYKAELDLGIPAALG
                     KYDIIFINVRTPYRHHHYQQCEDHAIHHSMLTRKAVDHLNKGGTCIALGYGTADRATE
                     NIISAVARSFRFSRVCQPKCAWENTEVAFVFFGKDNGNHLQDQDRLSVVLNNIYQGST
                     QHEAGRAPAYRVVRGDITKSNDEVIVNAANNKGQPGGGVCGALYRKWPGAFDKQPVAT
                     GKAHLVKHSPNVIHAVGPNFSRLSENEGDQKLSEVYMDIARIINNERFTKVSIPLLST
                     GIYAGGKDRVMQSLNHLFTAMDTTDADITIYCLDKQWESRIKEAITRKESVEELTEDD
                     RPVDIELVRVHPLSSLAGRPGYSTTEGKVYSYLEGTRFHQTAKDVAEINAMWPNKQEA
                     NEQICLYVLGESMNSIRSKCPVEESEASSPPHTIPCLCNYAMTAERVYRLRMAKNEQF
                     AVCSSFQLPKYRITGVQKIQCSKPVIFSGTVPPAIHPRKFASVTVEDTPVVQPERLVP
                     RRPAPPVPVPARIPSPPCTSTNGSTTSIQSLGEYQSASASSGAEISVDQVSLWSIPSA
                     TGFDVRTSSSLSLEQPTFPTMVVEAEIHASQGSLWSIPSITGFETRVPSPPSQDSRPS
                     TPSASASHTFVDLITFDSVAEILEDFSRSPFQFLSEIKPIPAPRTRVNNMSRSADTIK
                     PIPKPRKCQVKYTQPPGVARAISAAEFDEFVRRHSN"
     mat_peptide     47. .1642
                     /product="NS1"
     mat_peptide     1643. .4024
                     /product="NS2"
     mat_peptide     4025. .5680
                     /product="NS3"
     misc_feature    5681. .5683
                     /note="stop codon between NS3 and NS4 genes"
     CDS             5702. .7528
                     /note="unnamed protein product; Protein sequence is in
                     conflict with the conceptual translation"
                     /codon_start=1
                     /protein_id="CAA44847.1"
                     /db_xref="GI:1334981"
                     /db_xref="SPTREMBL:Q66581"
                     /translation="YIFSSETGQGHLQQKSTRQCKLQYPILERSVHEKFYAPRLDLER
                     EKLLQKKLQLCASEGNRSRYQSRKVENMKAITVERLLQGIGSYLSAEPQPVECYKVTY
                     PAPMYSSTASNSFSSAEVAVKVCNLVLQENFPTVASYNITDEYDAYLDMVDGASCCLD
                     TATFCPAKLRSFPKKHSYLRPEIRSAVPSPIQNTLQDVLAAATKRNCNVTQMRELPVL
                     DSAAFNVECFKKYACNDEYWDFYKTNPIRLTAENVTQYVTKLKGPKAAALFAKTHNLQ
                     PLHEIPMDRFVMDLKRDVKVTPGTKHTEERPKVQVIQAADPLATAYLCGIHRELVRRL
                     NAVLLPNIHTLFDMSAEDFDAIIAEHFQPGDAVLETDIASFDKSEDDAIAMSALMILE
                     DLGVDQALLNLIEAAFGNITSVHLPTGTRFKFGAMMKSGMFLTLFINTVVNIMIASRV
                     LRERLTTSPCAAFIGDDNIVKGVTSDALMAERCATWLNMEVKIIDAVVGVKAPYFCGG
                     FIVVDQITGTACRVADPLKRLFKLGKPLPLDDDQDVDRRRALHDEAARWNRIGITEEL
                     VKAVESRYEVNYVSLIITALTTLASSVSNFKHIRGHPITLYG"
     mat_peptide     5702. .7525
                     /product="NS4"
     misc_feature    7526. .7528
                     /note="stop codon between unstructural and structural
                     genes"
     misc_feature    7534
                     /note="start for 26S sub-RNA"
     CDS             7592. .11317
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAA44845.1"
                     /db_xref="GI:59186"
                     /db_xref="SPTREMBL:Q66579"
                     /translation="MFPYPTLNYPPMAPINPMAYRDPNPPRRRWRPFRPPLAAQIEDL
                     RRSIANLTLKQRAPNPPAGPPAKRKKPAPSLSLRRKKKRPPPPAKKQKRKPKPGKRQR
                     MCMKLESDKTFPIMLNGQVNGYACVVGGRVFKPLHVEGRIDNEQLAAIKLKKASIYDL
                     EYGDVPQCMKSDTLQYTSDKPPGFYNWHHGAVQYENNRFTVPRGVGGKGDSGRPILDN
                     KGRVVAIVLGGVNEGSRTALSVVTWNQKGVTVKDTPEGSEPWSLATVMCVLANITFPC
                     DQPPCMPCCYEKNPHETLTMLEQNYDSRAYDQLLDAAVKCNARRTRRDLDTHFTQYKL
                     ARPYIADCPNCGHSRCDSPIAIEEVRGDAHAGVIRIQTSAMFGLKTDGVDLAYMSFMN
                     GKTQKSIKIDNLHVRTSAPCSLVSHHGYYILAQCPPGDTVTVGFHDGPNRHTCTVAHK
                     VEFRPVGREKYRHPPEHGVELPCNRYTHKRADQGHYVEMHQPGLVADHSLLSIHSAKV
                     KITVPSGAQVKYYCKCPDVREGTTSSDYTTTCTDVKQCRAYLIDNKKWVYNSGRLPRG
                     EGDTFKGKLHVPFVPVKAKCIATLAPEPLVEHKHRTLILHLYPDHPTLLTTRSLGSDA
                     NPTRQWIERPTTVNFTVTGEGLEYTWGNHPPKRVWAQESGEGNPHGWPHEVVVYYYNR
                     YPLTTIIGLCTCVAIIMVSCVTSVWLLCRTRNLCITPYKLAPNAQVPILLALLCCIKP
                     TRADDTLQVLNYLWNNNQNFFWMQTLIPLAALIVCMRMLRCLFCCGPAFLLVCGALGA
                     AAYEHTAVMPNKVGIPYKALVERPGYAPVHLQIQLVNTRIIPSTNLEYITCKYKTKVP
                     SPVVKCCGATQCTSKPHPDYQCQVFTGVYPFMWGGAYCFCDTENTQMSEAYVERSEEC
                     SIDHAKAYKVHTGTVQAMVNITYGSVSWRSADVYVNGETPAKIGDAKLIIGPLSSAWS
                     PFDNKVVVHGHEVYNYDFPEYGTGKAGSFGDLQSRTSTSNDLYANTNLKLQRPQAGIV
                     HTPFTQAPSGFERWKRDKGAPLNDVAPFGCSIALEPLRAENCAVGSIPISIDIPDAAF
                     TRISETPTVSDLECKITECTYASDFGGIATVAYKSSKAGNCPIHSPSGVAVIKENDVT
                     LAESGSFTFHFSTANIHPAFKLQVCTSAVTCKGDCKPPKDHIVDYPAQHTESFTSAIS
                     ATAWSWLKVLVGGTSAFIVLGLIATAVVALVLFFHRH"
     mat_peptide     7592. .8371
                     /product="C protein"
     mat_peptide     8372. .8560
                     /product="E3 protein"
     mat_peptide     8561. .9820
                     /product="E2 protein"
     mat_peptide     9821. .9991
                     /product="6K protein"
     mat_peptide     9992. .11314
                     /product="E1 protein"
BASE COUNT     3322 a    2882 c    2805 g    2666 t
ORIGIN

  Query Match           54.9%;  Score 2745.2;  DB 14;  Length 11675;
  Best Local Similarity   72.3%;  Pred. No. 0;
  Matches 3592;  Conservative    0;  Mismatches 1368;  Indels    6;  Gaps    2;

Qy         14 ATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGT 73
              | | |||||| ||||| | | |||| |||||||||||| || |||||||| |  |||||||
Db         36 ACCTATCCAAAATGGAGAAAGTTCATGTTGACTTAGACGCAGACAGCCCATTCGTCAAGT 95

Qy         74 CGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACC 133
              |  | ||  |     |||||||| |||||||| ||||||| ||||||||||||||||||||
Db         96 CACTGCAAAGATGCTTTCCACATTTTGAGATAGAAGCAACGCAGGTCACTGACAATGACC 155

Qy        134 ATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACC 193
              |||| ||||| || |||||||||||  | || || |||||||||||| | ||||| ||
Db        156 ATGCTAATGCTAGGGCGTTTTCGCACCTAGCTACTAAGCTCATTGAGGGAGAAGTGGATA 215

Qy        194 GGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACC 253
                |||| || ||| |||| ||||| || ||||| || || ||| | |||||||| ||
```

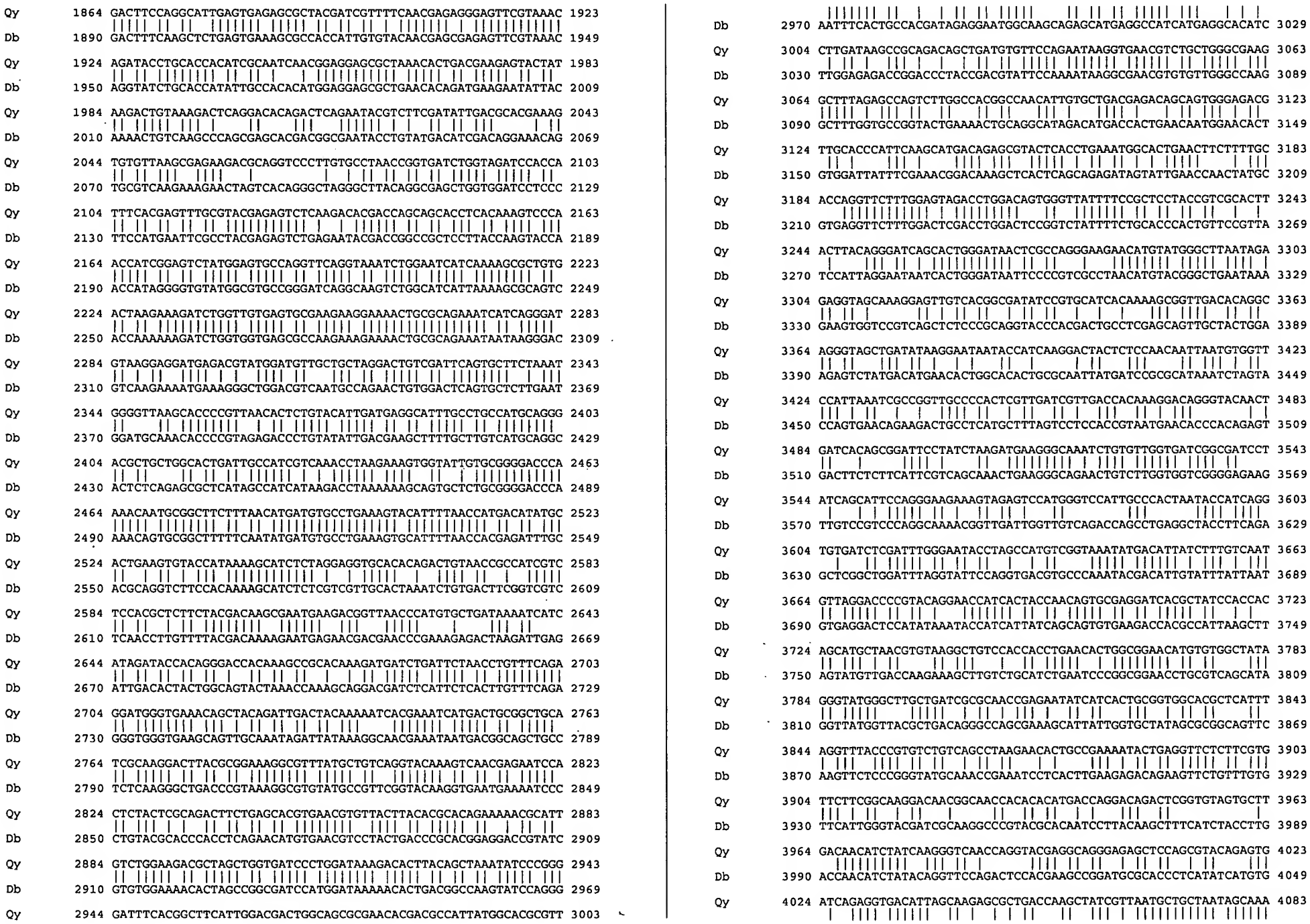

Db 216 CAGACCAGGTGATCCTGGATATTGGGAGCGCGCCTGTAAGGCATACGCATTCCAAACATA 275
Qy 254 GCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGGTATG 313
Db 276 AGTACCACTGTATTTGCCCAATGAAGAGCGCAGAAGACCCTGACAGACTCTACCGCTACG 335
Qy 314 CAGAAAGACTTAAGAAAAGTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACC 373
Db 336 CAGACAAGCTTAGAAAAAGTGATGTCACTGACAAATGTATTGCCTCTAAGGCCGCGGACC 395
Qy 374 TGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACG 433
Db 396 TGCTAACAGTAATGTCGACGCCTGACGCTGAGACACCCTCGTTATGCATGCACACTGACT 455
Qy 434 CCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCAC 493
Db 456 CAACTTGCAGGTACCACGGCTCCGTGGCCGTATATCAGGATGTATATGCAGTGCATGCAC 515
Qy 494 CGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTG 553
Db 516 CGACTTCCATTTACTACCAGGCGCTGAAAGGTGTACGAACTATCTATTGGATCGGGTTCG 575
Qy 554 ACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACT 613
Db 576 ATACTACACCGTTCATGTATAAGAACATGGCAGGCGCCTACCCTACATACAACACAAATT 635
Qy 614 GGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGG 673
Db 636 GGGCCGATGAAAGTGTGTTGGAAGCCAGAAATATAGGGCTGGGTAGTTCAGACTTGCACG 695
Qy 674 AGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGA 733
Db 696 AAAAGAGTTTCGGAAAAGTATCCATTATGAGGAAGAAGAAATTACAACCCACTAATAAAG 755
Qy 734 TCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGC 793
Db 756 TAATATTTTCTGTGGGGTCAACTATCTATACTGAAGAGAGAATACTGTTACGCAGCTGGC 815
Qy 794 ATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCA 853
Db 816 ATCTACCTAATGTCTTTCATCTAAAAGGTAAAACTAGCTTTACAGGCAGATGTAACACCA 875
Qy 854 TTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTA 913
Db 876 TCGTCAGCTGCGAAGGTTACGTTGTCAAGAAGATTACGCTCAGTCCTGGGATTTACGGGA 935
Qy 914 AAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAG 973
Db 936 AAGTGGATAATCTTGCTTCGACCATGCACCGAGAGGGATTCTTAAGTTGCAAGGTTACAG 995
Qy 974 ATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTT 1033
Db 996 ACACGTTAAGAGGGGAGAGGGTCTCTTTCCCCGTGTGTACGTACGTGCCAGCGACACTGT 1055
Qy 1034 GCGATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTAT 1093
Db 1056 GCGACCAGATGACCGGGATACTGGCGACTGACGTCAGTGTCGATGACGCCCAGAAGCTGC 1115
Qy 1094 TGGTTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAA 1153
Db 1116 TGGTTGGGCTCAACCAGCGAATTGTCGTCAATGGCAGAACACAACGTAACACAAATACCA 1175
Qy 1154 TGCAGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATC 1213
Db 1176 TGCAGAATTACCTATTACCAGTGGTAGCCCAGGCGTTCTCGCGGTG---CCGGGAATACC 1232
Qy 1214 GTGCCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCT 1273
Db 1233 GCGCAGACCTGGAAGACGAAAAAGGGCTAGGGGTACGGGAACGTTCCCTAGTCATGGGCT 1292
Qy 1274 GCTGCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCTGGTACGCAAA 1333
Db 1293 GCTGCTGGGCTTTCAAAACTCACAAGATCACATCCATTTACAAGAGACCTGGGACTCAAA 1352
Qy 1334 CAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCACG 1393
Db 1353 CTATCAAGAAGGTGCCCGCCGTATTTAATTCCTTTGTCATCCCACAACCAACCAGCTATG 1412
Qy 1394 GGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACTGTCAAACCCG 1453
Db 1413 GGCCTGATATAGGATTGCGTCGCCGAATTAAGATGCTATTCGACGCAAAGAAGGCACCCG 1472
Qy 1454 CACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAG 1513
Db 1473 CTCCAATTATTACTGAGGCCGACGTTGCACACCTTAAAGGCCTGCAGGATGAAGCTGAAG 1532
Qy 1514 AAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAA 1573
Db 1533 CCGTGGCTGAGGCTGAAGCCGTACGTGCAGCATTACCTCCACTTCTGCCGGAGGTCGATA 1592
Qy 1574 AAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGAGCAGGTAGCGTGG 1633
Db 1593 AGGAGACCGTAGAGGCTGACATCGACCTGATCATGCAGGAGGCAGGAGCAGGCAGCGTGG 1652
Qy 1634 AGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAGATTGGGTCTT 1693
Db 1653 AGACACCTAGACGACACATCAAGGTCACGACGTATCCAGGGGAAGAAATGATCGGCTCGT 1712
Qy 1694 ACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGTATCCACCCAT 1753
Db 1713 ACGCAGTGCTTTCACCACAAGCGGTCCTTAACAGCGAGAAGCTAGCTTGTATTCACCCGT 1772
Qy 1754 TGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATACAAAGTCGAGC 1813
Db 1773 TAGCTGAGCAAGTGCTCGTGATGACTCACAAAGGGCGCGCAGGACGATACAAGGTAGAGC 1832
Qy 1814 CATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAAGACTTCCAGG 1873
Db 1833 CATACCACGATAGAGTTATCGTCCCCAGTGGTACAGCTATACCAATCCCCGATTTCCAGG 1892
Qy 1874 CATTGAGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAACAGATACCTGC 1933
Db 1893 CTCTGAGTGAAAGTGCAACCATAGTATTTAACGAACGGGAGTTCGTTAACCGTTACTTAC 1952
Qy 1934 ACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTATAAGACTGTAA 1993
Db 1953 ACCACATTGCCGTTAACGGAGGGGCATTGAATACAGATGAAGAGTACTACAAGGTTGTGA 2012
Qy 1994 AGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAGTGTGTTAAGC 2053
Db 2013 AAAGCACTGAGACAGACTCTGAGTACGTATTTGACATCGACGCAAAGAAGTGCGTGAAGA 2072
Qy 2054 GAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCATTTCACGAGT 2113
Db 2073 AAGGGGATGCCGGACCAATGTGCCTGGTCGGCGAACTAGTTGACCCGCCATTCCACGAAT 2132
Qy 2114 TTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCAACCATCGGAG 2173
Db 2133 TTGCGTACGAGAGTTTAAAAACACGTCCTGCTGCACCACACAAAGTGCCTACCATCGGAG 2192
Qy 2174 TCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTGACTAAGAAAG 2233
Db 2193 TCTACGGAGTCCCAGGTTCTGGAAAGTCTGGTATAATCAAAAGCGCTGTTACCAAGCGTG 2252
Qy 2234 ATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGATGTAAGGAGGA 2293
Db 2253 ATCTGGTGGTCAGTGCAAAGAAAGAAAATTGCATGGAAATCATCAAAGACGTCAAACGCA 2312
Qy 2294 TGAGACGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAATGGGGTTAAGC 2353
Db 2313 TGCGCGGCATGGACATCGCCGCCCGCACAGTGGATTCAGTGCTGCTAAATGGGGTAAAAC 2372
Qy 2354 ACCCCGTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGGACGCTGCTGG 2413
Db 2373 ACTCCGTCGACACACTGTACATAGACGAGGCATTCGCTTGCCATGCAGGGACTCTGCTAG 2432



```
Db       4593 ATTCAACCACCGAGGGCAAGGTGTATTCGTACCTAGAGGGGACTAGGTTTCATCAAACTG 4652

Qy       4631 CCAAGGACATTGCCGAAATCCATGCAATGTGGCCCAACAAATCTGAGGCTAATGAGCAGA 4690

Db       4653 CCAAAGACGTAGCTGAGATTAACGCTATGTGGCCTAACAAGCAAGAAGCAAACGAGCAGA 4712

Qy       4691 TTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGCCCAGTAGAGG 4750

Db       4713 TTTGCTTATACGTGTTGGGAGAGAGTATGAACAGCATCCGCTCTAAGTGTCCTGTTGAAG 4772

Qy       4751 AGTCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTACGCTATGACGG 4810

Db       4773 AGTCGGAAGCCTCTTCCCCCCCTCACACCATCCCGTGTCTGTGCAACTATGCAATGACTG 4832

Qy       4811 CTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTATGCTCATCAT 4870

Db       4833 CAGAGCGAGTTTACAGATTACGTATGGCAAAGAACGAACAATTCGCAGTTTGTTCGTCCT 4892

Qy       4871 TCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGCAAACCAGTCC 4930

Db       4893 TTCAGTTACCGAAATACAGGATTACAGGGGTTCAGAAAATTCAATGCAGTAAACCTGTGA 4952

Qy       4931 TGTTTTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGC 4976

Db       4953 TATTCTCCGGCACTGTACCCCCGGCCATACATCCAAGAAAATTCGC 4998
```

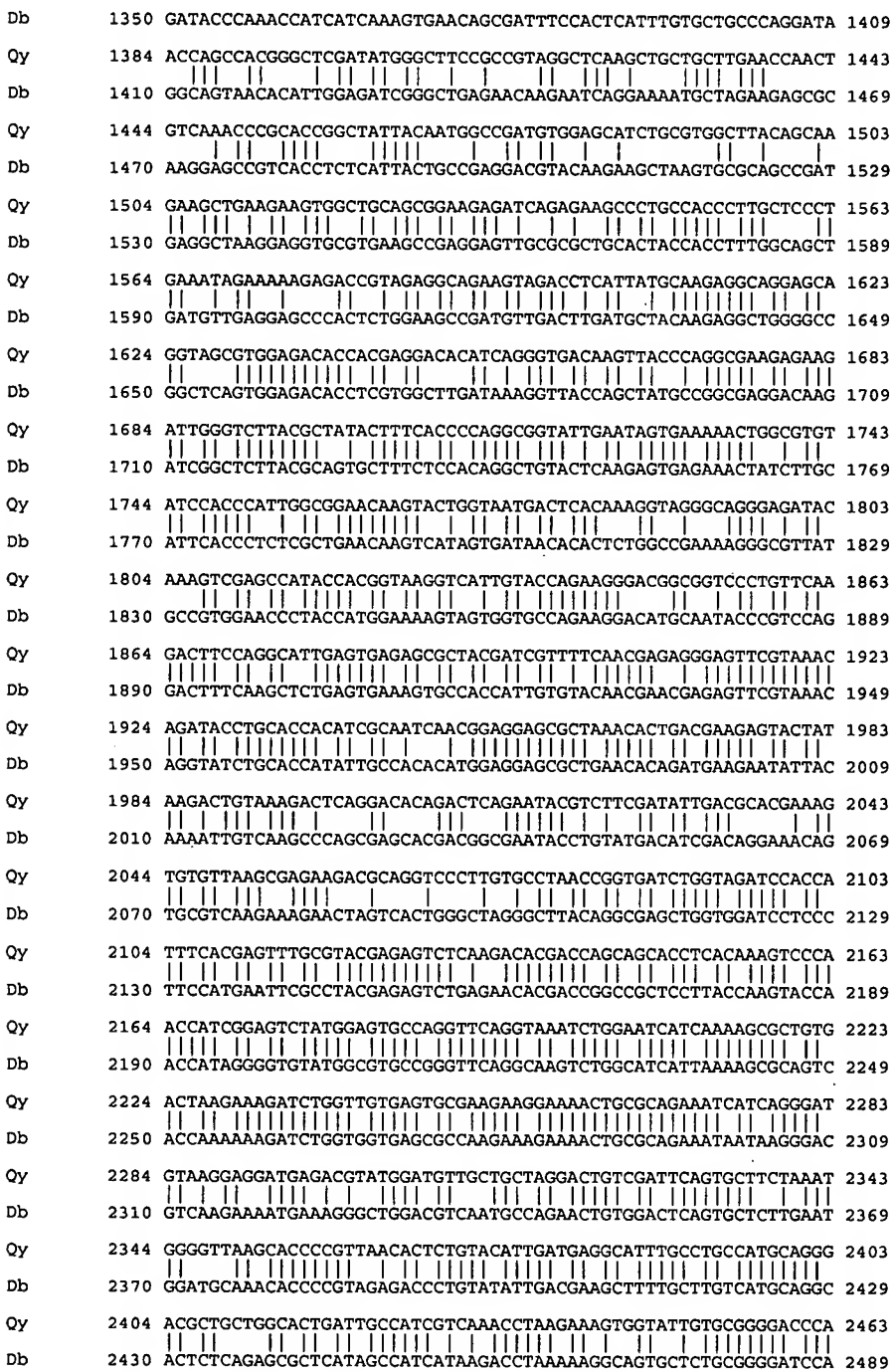
RESULT 6
VEU55362
LOCUS VEU55362 11441 bp RNA linear VRL 12-JUL-1999
DEFINITION Venezuelan equine encephalitis virus strain 83U434, complete genome.
ACCESSION U55362
VERSION U55362.2 GI:5442471
KEYWORDS .
SOURCE Venezuelan equine encephalitis virus
ORGANISM Venezuelan equine encephalitis virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae; Alphavirus; VEEV complex.
REFERENCE 1 (bases 8387 to 10063)
AUTHORS Weaver,S.C., Salas,R., Rico-Hesse,R., Ludwig,G.V., Oberste,M.S., Boshell,J. and Tesh,R.B.
TITLE Re-emergence of epidemic Venezuelan equine encephalomyelitis in South America. VEE Study Group
JOURNAL Lancet 348 (9025), 436-440 (1996)
MEDLINE 96332169
PUBMED 8709783
REFERENCE 2 (bases 1 to 11441)
AUTHORS Wang,E., Barrera,R., Boshell,J., Ferro,C., Freier,J.E., Navarro,J.C., Salas,R., Vasquez,C. and Weaver,S.C.
TITLE Genetic and phenotypic changes accompanying the emergence of epizootic subtype IC Venezuelan equine encephalitis viruses from an enzootic subtype ID progenitor
JOURNAL J. Virol. 73 (5), 4266-4271 (1999)
MEDLINE 99214366
PUBMED 10196323
REFERENCE 3 (bases 8387 to 10063)
AUTHORS Weaver,S.C.
TITLE Direct Submission
JOURNAL Submitted (18-APR-1996) Pathology, Univ. TX Med. Branch, 301 University Ave., Galveston, TX 77555-0609, USA
REFERENCE 4 (bases 8387 to 10063)
AUTHORS Weaver,S.C.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-1997) Pathology, Univ. TX Med. Branch, 301 University Ave., Galveston, TX 77555-0609, USA
REMARK Nucleotide and amino acid sequences updated by submitter
REFERENCE 5 (bases 1 to 11441)
AUTHORS Wang,E. and Weaver,S.C.
TITLE Direct Submission
JOURNAL Submitted (12-JUL-1999) Pathology, Univ. TX Med. Branch, 301 University Ave., Galveston, TX 77555-0609, USA
REMARK Sequence update by submitter
COMMENT On Jul 12, 1999 this sequence version replaced gi:3849808.
FEATURES Location/Qualifiers
source 1. .11441
/organism="Venezuelan equine encephalitis virus"
/mol_type="genomic RNA"
/strain="83U434"
/db_xref="taxon:11036"
CDS 45. .7523
/note="contains nonstructural proteins 1, 2, 3 and 4; in-frame termination codon is read through at low frequency"
/codon_start=1
/transl_except=(pos:5679. .5681,aa:OTHER)
/product="nonstructural polyprotein"
/protein_id="AAD43359.1"
/db_xref="GI:5442473"
/translation="MEKVHVDIEEDSPFLRALQRSFPQFEVEAKQVTDNDHANARAFS
HLASKLIETEVDPSDTILDIGSAPARRMYSKHKYHCICPMRCAEDPDRLYKYATKLKK
NCKEITDKELDKKMKELAAVMSDPDLETETMCLHDDESCRYEGQVAVYQDVYAVDGPT
SLYHQANKGVRVAYWIGFDTTPFMFKNLAGAYPSYSTNWADETVLTARNIGLCSSDVM
ERSRRGMSILRKKYLKPSNNVLFSVGSTIYHEKRDLLRSWHLPSVFHLRGKQNYTCRC
ETIVSCDGYVVKRIAISPGLYGKPSGYAATMHREGFLCCKVTDTLNGERVSFPVCTYV
PATLCDQMTGILATDVSADDAQKLLVGLNQRIVVNGRTQRNTNTMKNYLLPVVAQAFA
RWAKEYKEDQEDERPLGLRDRQLVMGCCWAFRRHKITSIYKRPDTQTIIKVNSDFHSF
VLPRIGSNTLEIGLRTRIRKMLEERKEPPPLITDEDVQEAKCAADEAKEVREAEELRA
ALPPLAADVEEPTLEADVDLMLQEAGAGSVETPRGLIKVTSYAGEDKIGSYAVLSPQA
VLKSEKLSCIHPLAEQVIVITHSGRKGRYAVEPYHGKVVVPEGHAIPVQDFQALSESA
TIVYNEREFVNRYLHHIATHGGALNTDEEYYKTVKPSEHDGEYLYDIDRKQCVKKELV
TGLGLTGELVDPPFHEFAYESLRIRPAAPYQVPTIGVYGVPGSGKSGIIKSAVTKKDL
VVSAKKENCAEIIRDVKKMKGLDVNARTVDSVLLNGCKHPVETLYIDEAFACHAGTLR
ALIAIIRPKKAVLCGDPKQCGFFNMMCLKVHFNHEICTQVFHKSISRRCTKSVTSVVS
TLFYDKRMRTTNPKETKIEIDTTGSTKPKQDDLILTCFRGWVKQLQIDYKGNEIMTAA
ASQGLTRKGVYAVRYKVNENPLYAPTSEHVNVLLTRTEDRIVWKTLAGDPWIKTLTAK
YPGNFTATIEEWQAEHEAIMRHILERPDPTDVFQNKANVCWAKALVPVLKTAGIDMTT
EQWNTVDYFETDKAHSAEIVLNQLCVRFFGLDLDSGLFSAPTVPLSIRNNHWDNSPSP
NMYGLNKEVVRQLSRRYPRLPRAVATGRVYDMNTGTLRNYDPRINLVPVNRRLPHALV
LHRNEHPQSDFSSFVSKLKGRTVLVVGEKLSVPGKTVDWLSDQPEATFRARLDLGIPG
DVPKYDIVFINVRTPYKYHHYQQCEDHAIKLSMLTKKACLHLNPGGTCVSIGYGYADR
ASESIIGAIARQFKFSRVCKPKSSLEETEVLFVFIGYDRKARTHNPYKLSSTLTNIYT
GSRLHEAGCAPSYHVVRGDIATATEGVIINAANSKGQPGGGVCGALYKKFPESFDLQP
IEVGKARLVKGAAKHIIHAVGPNFNKVSEVEGDKQLAEAYESIAKIVNDNNYKSVAIP
LLSTGIFSGNKDRLTQSLNHLLTALDTTDADVAIYCRDKKWEMTLKEAVARREAVEEI
CISDDSSVTEPDAELVRVHPKSSLAGRKGYSTSDGKTFSYLEGTKFHQAAKDIAEINA
MWPVATEANEQVCMYILGESMSSIRSKCPVEESEASTPPSTLPCLCIHAMTPERVQRL
KASRPEQITVCSSFPLPKYRITGVQKIQCSQPILFSPKVPAYIHPRKYLVETPPVEET
PEPLTENQSVEGTPEPAPVNVDATRTRTPEPIIIEEEEEDSISLLSDGPTHQVLQVEA
DIHGPPSVSSSSWSIPHASDFDVDSLSILDTLEGASVTSEAASAETNSYFARSMEFLA
RPVPAPRTVFRNPPHPAPRTRTPSLAPSRASSRTSLVSTPPGVNRVITREELEALTPS
RAPSRSASRTSLVSNPPGVNRVITREEFEAFVAQQQXRFDAGAYIFSSDTGQGHLQQK
SVRQTVLSEVVLERTELEISYAPRLDQEKEELLRKKLQLNPTPANRSRYQSRRVENMK
AMTARRILQGLGHYLKAEGKVECYRTLHPVPLYSSSVNRAFSSPKVAVEACNAMLKEN
FPTVASYCIIPEYDAYLDMVDGASCCLDTASFCPAKLRSFPKKHSYLEPTIRSAVPSA
IQNTLQNVLAAATKRNCNVTQMRELPVLDSAAFNVECFKKYACNNEYWETFKENPIRL
TEENVVNYITKLKGPKAAALFAKTHNLNMLQDIPMDRFVMDLKRDVKVTPGTKHTEER
PKVQVIQAADPLATAYLCGIHRELVRRLNAVLLPNIHTLFDMSAEDFDAIIAEHFQPG
DCVLETDIASFDKSEDDAMALTALMILEDLGVDAELLTLIEAAFGEISSIHLPTKTKF
KFGAMMKSGMFLTLFVNTVINIVIASRVLRERLTGSPCAAFIGDDNIVKGVKSDKLMA
DRCATWLNMEVKIIDAVVGEKAPYFCGGFILCDTVTGTACRVADPLKRLFKLGKPLAA
DDEHDDDRRRALHEESTRWNRVGILPELCKAVESRYETVGTSIIVMAMTTLASSVKSF
SYLRGAPITLYG"
CDS 7559. .11326
/note="contains capsid, E3, E2, 6K and E1 proteins"
/codon_start=1
/product="structural polyprotein"
/protein_id="AAC72000.2"
/db_xref="GI:5442472"
/translation="MFPFQPMYPMQPMPYRNPFAAPRRPWFPRTDPFLAMQVQELTRS
MANLTFKQRRDAPPEGPPAKKPKREAPQKQKGGGQGKKKKNQGKKKAKTGPPNPKAQS
GNKKKTNKKPGKRQRMVMKLESDKTFPIMLEGKINGYACVVGGKLFRPMHVEGKIDND
VLAALKTKKASKYDLEYADVPQNMRADTFKYTHEKPQGYYSWHHGAVQYENGRFTVPK
GVGAKGDSGRPILDNQGRVVAIVLGGVNEGSRTALSVVMWNEKGVTVKYTPENCEQWS
LVTTMCLLANVTFPCAQPPICYDRKPAETLAMLSVNVDNPGYDELLEAAVKCPGRKKR
STEELFKEYKLTRPYMARCIRCAVGSCHSPIAIEAVKSDGHDGYVRLQTSSQYGLDSS
```
```

```
                     GNLKGRTMRYDMHGTIEEIPLHQVSLHTSRPCHIVDGHGYFLLARCPAGDSITMEFKK
                     DSVTHSCSVPYEVKFNPVGRELYTHPPEHGAEQACQVYAHDAQNRGAYVEMHLPGSEV
                     DSSLVSLSGSSVTVTPPAGTSALVECECGGTKISETINTAKQFSQCTKKEQCRAYRLQ
                     NDKWVYNSDKLPKAAGATLKGKLHVPFLLADGKCTVPLAPEPMITFGFRSVSLKLHPK
                     NPTYLTTRQLADEPHYTHELISEPAVRNFTITEKGWEFVWGNHPPKRFWAQETAPGNP
                     HGLPHEVITHYYHRYPMSTILGLSICAAIVTVSVAASTWLFCKSRVSCLTPYRLTPNA
                     RMPLCLAVLCCARTARAETTWESLDHLWNNNQQMFWIQLLIPLAALIVVTRLLKCVCC
                     VVPFLVVAGAAGAGAYEHATTMPSQAGISYNTIVNRAGYAPLPISITPTKIKLIPTVN
                     LEYVTCHYKTGMDSPAIKCCGSQECTPTYRPDEQCKVFTGVYPFMWGGAYCFCDTENT
                     QVSKAYVMKSDDCLADHAEAYKAHTASVQAFLNITVGEHSIVTTVYVNGETPVNFNGV
                     KLTAGPLSTAWTPFDRKIVQYAGEIYNYDFPEYGAGQPGAFGDIQSRTVSSSDLYANT
                     NLVLQRPKAGAIHVPYTQAPSGFEQWKKDKAPSLKFTAPFGCEIYTNPIRAENCAVGS
                     IPLAFDIPDALFTRVSETPTLSAAECTLNECVYSSDFGGIATVKYSASKSGKCAVHVP
                     SGTATLKEAAVELTEQGSVTIHFSTANIHPEFRLQICTSYVTCKGDCHPPKDHIVTHP
                     QYHAQTFTAAVSKTAWTWLTSLLGGSAVIIIIGLVLATIVAMYVLTNQKHN"
BASE COUNT     3224 a    2825 c    2908 g    2484 t
ORIGIN

  Query Match           42.7%;  Score 2133.4;  DB 14;  Length 11441;
  Best Local Similarity  64.7%;  Pred. No. 0;
  Matches 3229;  Conservative    0;  Mismatches 1736;  Indels   24;  Gaps    3;

Qy         10 ACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTC 69
Db         30 AATTACCTACCCAAAATGGAGAAAGTTCACGTTGACATCGAGGAAGACAGCCCATTCCTC 89

Qy         70 AAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAAT 129
Db         90 AGAGCTTTGCAACGGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCACTGATAAT 149

Qy        130 GACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTC 189
Db        150 GACCATGCTAATGCCAGAGCGTTTTCGCATCTGGCTTCAAAACTGATCGAAACGGAGGTG 209

Qy        190 GACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAAT 249
Db        210 GACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCCGCTCGCAGAATGTATTCTAAG 269

Qy        250 CACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGG 309
Db        270 CATAAGTATCATTGCATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGATTGTACAAG 329

Qy        310 TATGCAGAAAGACTTAAGAAAAGT------GACATTACCGACAAGAACATAGCCTCTAAG 363
Db        330 TATGCAACTAAGCTGAAGAAAAATTGCAAGGAAATAACTGACAAGGAATTGGACAAGAAA 389

Qy        364 GCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATG 423
Db        390 ATGAAGGAGCTCGCCGCCGTCATGAGCGACCCTGACCTGGAAACTGAGACTATGTGCCTC 449

Qy        424 CACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCA 483
Db        450 CACGACGACGAGTCGTGTCGCTACGAGGGGCAAGTCGCTGTTTACCAGGATGTATACGCA 509

Qy        484 GTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGG 543
Db        510 GTTGACGGACCGACAAGTCTCTATCACCAAGCCAACAAGGGAGTTAGAGTCGCCTACTGG 569

Qy        544 ATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTAC 603
Db        570 ATAGGCTTTGACACCACCCCTTTTATGTTTAAGAACTTGGCTGGAGCATATCCATCATAC 629

Qy        604 AACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCA 663
Db        630 TCTACCAACTGGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCCTATGCAGCTCC 689

Qy        664 GATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCT 723
Db        690 GACGTCATGGAGCGGTCACGTAGAGGGATGTCCATTCTTAGGAAGAAGTATTTGAAACCA 749

Qy        724 ACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTA 783
Db        750 TCCAATAATGTCCTATTCTCTGTTGGCTCGACCATCTACCACGAGAAGAGGGATTTACTG 809

Qy        784 CGTAGCTGGCATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGA 843
Db        810 AGGAGCTGGCACCTGCCGTCTGTATTTCACTTACGTGGCAAGCAAAATTACACATGTCGG 869

Qy        844 TGTGGGACCATTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGA 903
Db        870 TGTGAGACTATAGTTAGTTGCGACGGGTACGTCGTTAAAAGAATAGCTATCAGTCCAGGC 929

Qy        904 CTATACGGTAAAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGC 963
Db        930 CTGTATGGGAAGCCTTCAGGCTATGCTGCCACGATGCACCGCGAGGGATTCTTGTGCTGC 989

Qy        964 AAAGTCACAGATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCA 1023
Db        990 AAAGTGACAGACACATTGAACGGGGAGAGGGTCTCTTTTCCCGTGTGCACGTATGTGCCA 1049

Qy       1024 GCCACACTTTGCGATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCA 1083
Db       1050 GCTACATTGTGTGACCAAATGACTGGCATACTGGCAACAGATGTCAGTGCGGACGACGCG 1109

Qy       1084 CAAAAACTATTGGTTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAAT 1143
Db       1110 CAAAAACTGCTGGTTGGGCTCAACCAGCGCATAGTCGTCAACGGTCGCACCCAAAGAAAC 1169

Qy       1144 ACTAACACAATGCAGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCG 1203
Db       1170 ACCAATACCATGAAGAATTATCTTTGCCCGTAGTGGCCCAGGCATTTGCTAGGTGGGCA 1229

Qy       1204 CGTGAACATCGTGCCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTT 1263
Db       1230 AAAGAATATAAGGAAGATCAAGAAGATGAGAGGCCACTAGGACTACGAGATAGACAGTTA 1289

Qy       1264 ACTATGGGCTGCTGCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCT 1323
Db       1290 GTTATGGGGTGCTGCTGGGCTTTTAGAAGGCACAAGATAACATCTATTTATAAGCGCCCA 1349

Qy       1324 GGTACGCAAACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTT 1383
Db       1350 GATACCCAAACCATCATCAAAGTGAACAGCGATTTCCACTCATTTGTGCTGCCCAGGATA 1409

Qy       1384 ACCAGCCACGGGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACT 1443
Db       1410 GGCAGTAACACATTGGAGATCGGGCTGAGAACAAGAATCAGGAAAATGCTAGAAGAGCGC 1469

Qy       1444 GTCAAACCCGCACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAA 1503
Db       1470 AAGGAGCCGCCACCTCTCATTACTGACGAGGACGTACAAGAAGCTAAGTGCGCAGCCGAT 1529

Qy       1504 GAAGCTGAAGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCT 1563
Db       1530 GAGGCTAAGGAGGTGCGTGAAGCCGAGGAGTTGCGCGCTGCACTACCACCTTTGGCAGCT 1589

Qy       1564 GAAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGAGCA 1623
Db       1590 GATGTTGAGGAGCCCACTCTGGAAGCCGATGTTGACTTGATGCTACAAGAGGCTGGGGCC 1649

Qy       1624 GGTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAG 1683
Db       1650 GGCTCAGTGGAGACACCTCGTGGCTTGATAAAGGTTACCAGCTATGCCGGCGAGGACAAG 1709

Qy       1684 ATTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGT 1743
Db       1710 ATCGGCTCTTACGCAGTGCTTTCTCCACAGGCTGTACTCAAGAGTGAGAAACTATCTTGC 1769

Qy       1744 ATCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATAC 1803
Db       1770 ATTCACCCTCTCGCTGAACAAGTCATAGTGATAACACACTCTGGCCGAAAAGGGCGTTAT 1829

Qy       1804 AAAGTCGAGCCATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAA 1863
Db       1830 GCCGTGGAACCCTACCATGGAAAAGTAGTGGTGCCAGAGGGACATGCAATACCCGTCCAG 1889
```



```
Db       4050 GTGCGAGGGGACATTGCCACGGCCACCGAAGGAGTGATCATAAATGCTGCTAACAGCAAA 4109

Qy       4084 GGTCAACCAGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTTGAT 4143

Db       4110 GGACAACCTGGCGGAGGGGTGTGCGGAGCGCTGTATAAGAAATTCCCGGAAAGCTTCGAT 4169

Qy       4144 AGACAGCCAATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACC---GCTCATCATA 4200

Db       4170 TTACAGCCGATCGAAGTGGGAAAAGCGCGACTGGTCAAAGGTGCAGCTAAACATATCATT 4229

Qy       4201 CATGCTGTAGGACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTAAGCTCGCA 4260

Db       4230 CATGCCGTAGGGCCCAACTTCAACAAAGTTTCGGAAGTTGAAGGTGACAAACAGTTGGCA 4289

Qy       4261 GCTGCCTACATGAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTA 4320

Db       4290 GAGGCTTATGAGTCCATCGCTAAAATTGTCAACGATAACAATTACAAGTCAGTAGCGATT 4349

Qy       4321 CCGCTACTGTCAACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCAT 4380

Db       4350 CCACTGTTGTCCACCGGCATCTTTTCCGGGAACAAAGACCGACTAACCCAATCATTGAAC 4409

Qy       4381 CACCTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAA 4440

Db       4410 CATTTGCTGACAGCTTTAGACACCACTGATGCAGATGTAGCCATATACTGCAGGGACAAG 4469

Qy       4441 CAATGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATTCTGGA- 4499

Db       4470 AAATGGGAAATGACTCTCAAGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGGAGATATGC 4529

Qy       4500 --------------TGATGACAAGCCAGTAGACATTGACTTGGTCAGGGTCCACCCAAAC 4545

Db       4530 ATATCCGACGATTCTTCGGTGACAGAACCGGATGCAGAGCTGGTGAGGGTACATCCGAAG 4589

Qy       4546 AGCTCTTTGGCAGGCAGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTG 4605

Db       4590 AGCTCTTTGGCTGGAAGGAAGGGCTACAGCACAAGTGATGGCAAGACTTTCTCATATTTG 4649

Qy       4606 GAAGGTACACGATTCCATCAGACCGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCC 4665

Db       4650 GAAGGGACCAAATTTCACCAGGCGGCCAAGGATATAGCAGAAATTAATGCCATGTGGCCC 4709

Qy       4666 AACAAATCTGAGGCTAATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGC 4725

Db       4710 GTTGCAACGGAGGCCAATGAGCAAGTATGCATGTATATCCTTGGTGAAAGCATGAGCAGC 4769

Qy       4726 ATCCGCTCCAAATGCCCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCA 4785

Db       4770 ATTAGGTCAAAATGCCCCGTCGAAGAGTCGGAAGCCTCCACACCACCTAGCACGCTGCCT 4829

Qy       4786 TGCCTGTGTAATTACGCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAA 4845

Db       4830 TGCTTGTGCATCCATGCTATGACTCCAGAAAGAGTACAACGCCTAAAAGCCTCACGTCCA 4889

Qy       4846 GAACAGTTCGCCGTATGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAG 4905

Db       4890 GAACAAATTACTGTGTGCTCATCCTTTCCATTGCCGAAGTATAGAATCACTGGTGTGCAG 4949

Qy       4906 AAGCTACAGTGCAGCAAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCC 4965

Db       4950 AAGATCCAGTGCTCCCAGCCTATACTGTTCTCACCTAAGGTGCCTGCGTACATTCATCCA 5009

Qy       4966 AGGAAGTAC 4974

Db       5010 CGGAAGTAC 5018


RESULT 7
AF004472
LOCUS       AF004472               11420 bp    RNA     linear   VRL 12-JUL-1999
DEFINITION  Venezuelan equine encephalitis virus strain 66457, complete genome.
ACCESSION   AF004472
VERSION     AF004472.2  GI:5442461
KEYWORDS    .
SOURCE      Venezuelan equine encephalitis virus
  ORGANISM  Venezuelan equine encephalitis virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
            Alphavirus; VEEV complex.
REFERENCE   1  (bases 8366 to 10039)
  AUTHORS   Powers,A.M., Oberste,M.S., Brault,A.C., Rico-Hesse,R.,
            Schmura,S.M., Smith,J.F., Kang,W., Sweeney,W.P. and Weaver,S.C.
  TITLE     Repeated emergence of epidemic/epizootic Venezuelan equine
            encephalitis from a single genotype of enzootic subtype ID virus
  JOURNAL   J. Virol. 71 (9), 6697-6705 (1997)
  MEDLINE   97404681
   PUBMED   9261393
REFERENCE   2  (bases 1 to 11420)
  AUTHORS   Wang,E., Barrera,R., Boshell,J., Ferro,C., Freier,J.E.,
            Navarro,J.C., Salas,R., Vasquez,C. and Weaver,S.C.
  TITLE     Genetic and phenotypic changes accompanying the emergence of
            epizootic subtype IC Venezuelan equine encephalitis viruses from an
            enzootic subtype ID progenitor
  JOURNAL   J. Virol. 73 (5), 4266-4271 (1999)
  MEDLINE   99214366
   PUBMED   10196323
REFERENCE   3  (bases 8366 to 10039)
  AUTHORS   Powers,A.M., Oberste,M.S., Brault,A.C., Rico-Hesse,R.,
            Schmura,S.M., Smith,J.F., Kang,W., Sweeney,W.P. and Weaver,S.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-1997) Pathology, Univ. Texas Med. Branch,
            Galveston, TX 77555-0609, USA
REFERENCE   4  (bases 1 to 11420)
  AUTHORS   Wang,E. and Weaver,S.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-1999) Pathology, Univ. Texas Med. Branch,
            Galveston, TX 77555-0609, USA
  REMARK    Sequence update by submitter
COMMENT     On Jul 12, 1999 this sequence version replaced gi:3849802.
FEATURES             Location/Qualifiers
     source          1. .11420
                     /organism="Venezuelan equine encephalitis virus"
                     /mol_type="genomic RNA"
                     /strain="66457"
                     /db_xref="taxon:11036"
     CDS             45. .7502
                     /note="contains nonstructural proteins 1, 2, 3 and 4;
                     in-frame termination codon is read through at low
                     frequency"
                     /codon_start=1
                     /transl_except=(pos:5658. .5660,aa:OTHER)
                     /product="nonstructural polyprotein"
                     /protein_id="AAD43356.1"
                     /db_xref="GI:5442463"
                     /translation="MEKVHVDIEEDSPFLRALQRSFPQFEVEAKQVTDNDHANARAFS
                     HLASKLIETEVDPSDTILDIGSAPARRMYSKHKYHCICPMRCAEDPDRLYKYATKLKK
                     NFKEITDKELDKKMKELAAVMSDPDLETETMCLHDDESCRYEGQVAVYQDVYAVDGPT
                     SLYHQANKGVRVAYWIGFDTTPFMFKNLAGAYPSYSTNWADETVLTARNIGLCSSDVM
                     ERSRRGMSILRKKYLKPSNNVLFSVGSTIYHEKRDLLRSWHLPSVFHLRGKQNYTCRC
                     ETIVSCDGYVVKRIAISPGLYGKPSGYAATMHREGFLCCKVTDTLNGERVSFPVCTYV
                     PATLCDQMTGILATDVSADDAQKLLVGLNQRIVVNGRTQRNTNTMKNYLLPVVAQAFA
                     RWAKEYKEDQEDERPLGLRDRQLVMGCCWAFRRHKITSIYKRPDTQTIIKVNSDFHSF
                     VLPRIGSNTLEIGLRTRIRKMLEERKEPSPLITAEDVQEAKCAADEAKEVREAEELRA
                     ALPPLAADVEEPTLEADVDLMLQEAGAGSVETPRGLIKVTSYAGEDKIGSYAVLSPQA
                     VLKSEKLSCIHPLAEQVIVITHSGRKGRYAVEPYHGKVVVPEGHAIPVQDFQALSESA
                     TIVYNEREFVNRYLHHIATHGGALNTDEEYYKIVKPSEHDGEYLYDIDRKQCVKKELV
                     TGLGLTGELVDPPFHEFAYESLRTRPAAPYQVPTIGVYGVPGSGKSGIIKSAVTKKDL
                     VVSAKKENCAEIIRDVKKMKGLDVNARTVDSVLLNGCKHPVETLYIDEAFACHAGTLR
                     ALIAIIRPKKAVLCGDPKQCGFFNMMCLKVHFNHEICTQVFHKSISRRCTKSVTSVVS
                     TLFYDKRMRTTNPKETKIEIDTTGSTKPEQDDLILTCFRGWVKQLQIDYKGNEIMTAA
                     ASQGLTRKGVYAVRYKVNENPLYAPTSEHVNVLLTRTEDRIVWKTLAGDPWIKTLTAK
                     YPGNFTATIEEWQAEHEAIMRHILERPDPTDAFQNKANVCWAKALVPVLKTAGIDMTT
                     EQWNTVDYFETDKAHSAEIVLNQLCVRFFGLDLDSGLFSAPTVPLSIRNNHWDNSPSP
                     NMYGLNKEVVRQLSRRYPRLPRAVATGRVYDMNTGTLCNYDPRINLVPVNRRLPHALV
                     LHHNEHPQSDFSSFVSKLKGRTVLVVGEKLSVPGKTVDWLSDQPEATFRARLDLGIPG
                     AVPKYDIVFINVRTPYKYHHYQQCEDHAIKLSMLTKKACLHLNPGGTCVSIGYGYADR
                     ASESIIGAIARQFKFSRVCKPKSSLEETEVLFVFIGYDRKARTHNPYKLSSTLTNIYT
```

```
                     GSRLHEAGCAPSYHVVRGDIATATEGVIINAANSKGQPGGGVCGALYKKFPESFDLQP
                     IEVGKARLVKGAAKHIIHAVGPNFNKVSEVEGDKQLAEAYESIAKIVNDNNYKSVAIP
                     LLSTGIFSGNKDRLTQSLNHLLTALDTTDADVAIYCRDKKWEMTLKEAVARREAVEEI
                     CISDDSSVTEPDAELVRVHPKSSLAGRKGYSTSDGKTFSYLEGTKFHQAAKDIAEINA
                     MWPVATEANEQVCMYILGESMSSIRSKCPVEESEASTPPSTLPCLCIHAMTPERVQRL
                     KASRPEQITVCSSFPLPKYRITGVQKIQCSQPILFSPKVPAYIHPRKYLVETPPVEEN
                     QSTEGTPEPAPVNVDATRTRTPEPIIIEEEEEDSISLLSDGPTHQVLQVEADIHGPPS
                     VSSSSWSIPHASDFDVDSLSILDTLEGASVTSEAASAETNSYFARSMEFLARPVPAPR
                     TVFRNPPHPAPRTRTPSLAPSRASSRTSLVSTPPSANRVITREELEALTPSRAPSRSA
                     SRTSLVSNPPGVNRVITREEFEAFVAQQQXRFDAGAYIFSSDTGQGHLQQKSVRQTVL
                     SEVVLERTELEISYAPRLDQEKEELLRKKLQLNPTPANRSRYQSRRVENMKAITARRI
                     LQGLGHYLKAEGKVECYRTLHPVPLYSSSVNRAFSSPKVAVEACNAMLKENFPTVASY
                     CIIPEYDAYLDMVDGASCCLDTASFCPAKLRSFPKKHSYLEPTIRSAVPSAIQNTLQN
                     VLAAATKRNCNVTQMRELPVLDSAAFNVECFKKYACNNEYWETFKENPIRLTEENVVN
                     YITKLKGPKAAALFAKTHNLNMLQDIPMDRFVMDLKRDVKVTPGTKHTEERPKVQVIQ
                     AADPLATAYLCGIHRELVRRLNAVLLPNIHTLFDMSAEDFDAIIAEHFQPGDCVLETD
                     IASFDKSEDDAMALTALMILEDLGVDAELLTLIEAAFGEISSIHLPTKTKFKFGAMMK
                     SGMFLTLFVNTVINIVIASRVLRERLTGSPCAAFIGDDNIVKGVKSDKLMADRCATWL
                     NMEVKIIDAVVGEKAPYFCGGFILCDTVTGTACRVADPLKRLFKLGKPLAADDEHDDD
                     RRRALHEESTRWNRVGILPELCKAVESRYETVGTSIIVMAMTTLASSVKSFSYLRGAP
                     ITLYG"
     CDS             7538. .11305
                     /note="contains capsid, E3, E2, 6K and E1 proteins"
                     /codon_start=1
                     /product="structural polyprotein"
                     /protein_id="AAC71997.2"
                     /db_xref="GI:5442462"
                     /translation="MFPFQPMYPMQPMPYRNPFAAPRRPWFPRTDPFLAMQVQELTRS
                     MANLTFKQRRDAPPEGPPAKKPKREAPQKQKGGGQGKKKKNQGKKKAKTGPPNPKVQS
                     GNKKKTNKKPGKRQRMVMKLESDKTFPIMLEGKINGYACVVGGKLFRPMHVEGKIDND
                     VLAALKTKKASKYDLEYADVPQNMRADTFKYTHEKPQGYYSWHHGAVQYENGRFTVPK
                     GVGAKGDSGRPILDNQGRVVAIVLGGVNEGSRTALSVVMWNEKGVTVKYTPENCEQWS
                     LVTTMCLLANVTFPCAQPPICYDRKPAETLAMLSVNVDNSGYDELLEAAVKCPGRKRR
                     STEELFKEYKLTRPYMARCIRCAVGSCHSPIAIEAVKSDGHDGYVRLQTSSQYGLDSS
                     GNLKGRTMRYDMHGTIEEIPLHQVSLHTSRPCHIVDGHGYFLLARCPAGDSITMEFKK
                     NLVTHSCSVPYEVKFNPVGRELYTHPPEHGAEQACQVYAHDAQNRGAYVEMHLPGSEV
                     DSSLVSLSGSSVTVTPPAGTSALVECECGGTKISETINTAKQFSQCTKKEQCRAYRLQ
                     NDKWVYNSDKLPKAAGATLKGKLHVPFLLADGKCTVPLAPEPMITFGFRSVSLKLHPK
                     NPTYLTTRQLADEPHYTHELISEPAVRNFTVTEKGWEFVWGNHPPKRFWAQETAPGNP
                     HGLPHEVITHYYHRYPMSTILGLSICAAIVTVSVAASTWLFCKSRVSCLTPYRLTPNA
                     RMPLCLAVLCCARTARAETIWESLDHLWNNNQQMFWIQLLIPLAALIVVTRLLKCVCC
                     VVPFLVVAGAAGAGAYEHATTMPSQAGISYNTIVNRAGYAPLPISITPTKIKLIPTVN
                     LEYVTCHYKTGMDSPAIKCCGSQECTPTYRPDEQCKVFTGVYPFMWGGAYCFCDTENT
                     QVSKAYVMKSDDCLADHAEAYKAHTASVQAFLNITVGEHSIVTTVYVNGETPVNFNGV
                     KLTAGPLSTAWTPFDRKIVQYAGEIYNYDFPEYGAGQPGAFGDIQSRTVSSSDLYANT
                     NLVLQRPKAGAIHVPYTQAPSGFEQWKKDKAPSLKFTAPFGCEIYTNPIRAENCAVGS
                     IPLAFDIPDALFTRVSETPTLSAAECTLNECVYSSDFGGIATVKYSASKSGKCAVHVP
                     SGTATLKEAAVELTEQGSVTIHFSTANIHPEFRLQICTSYVTCKGDCHPPKDHIVTHP
                     QYHAQTFTAAVSKTAWTWLTSLLGGSAVIIIIGLVLATIVAMYVLTNQKHN"
BASE COUNT     3215 a   2820 c   2900 g   2485 t
ORIGIN

  Query Match           42.6%;  Score 2128.6;  DB 14;  Length 11420;
  Best Local Similarity  64.7%;  Pred. No. 0;
  Matches 3226;  Conservative    0;  Mismatches 1739;  Indels   24;  Gaps    3;

Qy         10 ACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTC 69
              | | | | | |||| ||||| | | ||||||||||| | || |  |||||||| |   ||
Db         30 AATTACCTACCCAAAATGGAGAAAGTTCACGTTGACATCGAGGAAGACAGCCCATTCCTC 89

Qy         70 AAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAAT 129
              |    | || || ||||  || || || |||||| | ||||| | ||||||||||||| |||
Db         90 AGAGCTTTGCAACGGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCACTGATAAT 149

Qy        130 GACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTC 189
              |||||||| |||||||||||||||||||||| ||||  |||| || || || |  || ||
Db        150 GACCATGCTAATGCCAGAGCGTTTTCGCATCTGGCTTCAAAACTGATCGAAACGGAGGTG 209

Qy        190 GACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAAT 249
              ||||    | |    ||| | ||||||||||||||||||| | |         |||| ||
Db        210 GACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCCGCCCGCAGAATGTATTCTAAG 269

Qy        250 CACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGG 309
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        270 CATAAGTATCATTGCATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGATTGTACAAG 329

Qy        310 TATGCAGAAAGACTTAAGAAAAGT------GACATTACCGACAAGAACATAGCCTCTAAG 363
              ||||||    | || |||||| |       || || || |||||| | | | |   ||
Db        330 TATGCAACTAAGCTGAAGAAAAATTTCAAGGAAATAACTGACAAGGAATTGGACAAGAAA 389

Qy        364 GCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATG 423
              |   || ||    | ||||||      || |||   || |||    || |||| |
Db        390 ATGAAGGAGCTCGCCGCCGTCATGAGCGACCCTGACCTGGAAACTGAGACTATGTGCCTC 449

Qy        424 CACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCA 483
              |||   ||||   ||||| | |||   ||   || || || ||||| ||||| ||||||
Db        450 CACGACGACGAGTCGTGTCGCTACGAAGGGCAAGTCGCTGTTTACCAGGATGTATACGCA 509

Qy        484 GTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGG 543
              ||  | | |||||||     |||| ||||| ||    || |||||||        ||||||
Db        510 GTTGACGGACCGACAAGTCTCTATCACCAAGCCAACAAGGGAGTTAGAGTCGCCTACTGG 569

Qy        544 ATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTAC 603
              |||||||||||||| ||||||||||||   || ||| |||| ||  | || || | |||
Db        570 ATAGGCTTTGACACCACCCCTTTTATGTTTAAGAACTTGGCTGGAGCATATCCATCATAC 629

Qy        604 AACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCA 663
                 || |||||||| |||||  |  || ||   || |||||||| |||||  | |  ||
Db        630 TCTACCAACTGGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCCTATGCAGTTCC 689

Qy        664 GATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCT 723
              ||  |   |||| |    | ||    | || || |||||||||||    | ||||
Db        690 GACGTCATGGAGCGGTCACGTAGAGGGATGTCCATTCTTAGGAAGAAGTATTTGAAACCA 749

Qy        724 ACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTA 783
              | |||||  || |||||| ||||| || || ||||||   ||  | ||     | |
Db        750 TCCAATAATGTCCTATTCTCTGTTGGCTCGACCATCTACCACGAGAAGAGGGACTTACTG 809

Qy        784 CGTAGCTGGCATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGA 843
              | |||||||| || ||   || || |||||   || |||   || | |||| || |
Db        810 AGGAGCTGGCACCTGCCGTCTGTATTTCACTTACGTGGCAAGCAAAATTACACATGTCGG 869

Qy        844 TGTGGGACCATTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGA 903
              |||| ||| || || || || || ||||||||| | |||| ||| | ||||| ||||
Db        870 TGTGAGACTATAGTTAGTTGCGACGGGTACGTCGTTAAAAGAATAGCTATCAGTCCAGGC 929

Qy        904 CTATACGGTAAAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGC 963
              || || || ||          ||  ||  |||| ||||| |||||||| |||||| | |||
Db        930 CTGTATGGGAAGCCATCAGGCTATGCTGCCACGATGCACCGCGAGGGATTCTTGTGCTGC 989

Qy        964 AAAGTCACAGATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCA 1023
              ||||| ||||| ||  ||  ||| |||||||| |||||| | ||||| |||||||| |||
Db        990 AAAGTGACAGACACATTGAACGGGGAGAGGGTCTCTTTTCCCGTGTGCACGTATGTGCCA 1049

Qy       1024 GCCACACTTTGCGATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCA 1083
              || ||| | || || || ||||| || || |||||||| || || |||| ||| ||||||
Db       1050 GCTACATTGTGTGACCAAATGACTGGCATACTGGCAACAGATGTCAGTGCGGACGACGCG 1109

Qy       1084 CAAAAACTATTGGTTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAAT 1143
              |||||||| |||||||||||||| | || |||||||| ||| | || ||||||||
Db       1110 CAAAAACTGCTGGTTGGGCTCAACCAGCGCATAGTCGTCAACGGTCGCACCCAAAGAAAC 1169

Qy       1144 ACTAACACAATGCAGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCG 1203
              || || || ||| |||||||||| || || || || |||||||| ||| | |||||||
Db       1170 ACCAATACCATGAAGAACTATCTTTTGCCCGTAGTGGCCCAGGCATTTGCTAGGTGGGCA 1229

Qy       1204 CGTGAACATCGTGCCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTT 1263
              ||| ||   |  ||    || || |||||   |||||  | || ||   |      |
Db       1230 AAAGAATATAAGGAAGATCAAGAAGATGAGAGGCCACTAGGACTACGAGATAGACAGTTA 1289

Qy       1264 ACTATGGGCTGCTGCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCT 1323
              ||||||  |||||||||||||||  |  | || || || ||||| || || |||   ||
Db       1290 GTTATGGGGTGCTGCTGGGCTTTTAGAAGGCACAAGATAACATCTATTTATAAGCGCCCA 1349

Qy       1324 GGTACGCAAACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTT 1383
              | ||| ||||| || |  |||||      ||  ||  |||||||||| | ||  |  |
```


Db 1350 GATACCCAAACCATCATCAAAGTGAACAGCGATTTCCACTCATTTGTGCTGCCCAGGATA 1409
Qy 1384 ACCAGCCACGGGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACT 1443
Db 1410 GGCAGTAACACATTGGAGATCGGGCTGAGAACAAGAATCAGGAAAATGCTAGAAGAGCGC 1469
Qy 1444 GTCAAACCCGCACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAA 1503
Db 1470 AAGGAGCCGTCACCTCTCATTACTGCCGAGGACGTACAAGAAGCTAAGTGCGCAGCCGAT 1529
Qy 1504 GAAGCTGAAGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCT 1563
Db 1530 GAGGCTAAGGAGGTGCGTGAAGCCGAGGAGTTGCGCGCTGCACTACCACCTTTGGCAGCT 1589
Qy 1564 GAAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGAGCA 1623
Db 1590 GATGTTGAGGAGCCCACTCTGGAAGCCGATGTTGACTTGATGCTACAAGAGGCTGGGGCC 1649
Qy 1624 GGTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAG 1683
Db 1650 GGCTCAGTGGAGACACCTCGTGGCTTGATAAAGGTTACCAGCTATGCCGGCGAGGACAAG 1709
Qy 1684 ATTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGT 1743
Db 1710 ATCGGCTCTTACGCAGTGCTTTCTCCACAGGCTGTACTCAAGAGTGAGAAACTATCTTGC 1769
Qy 1744 ATCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATAC 1803
Db 1770 ATTCACCCTCTCGCTGAACAAGTCATAGTGATAACACACTCTGGCCGAAAAGGGCGTTAT 1829
Qy 1804 AAAGTCGAGCCATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAA 1863
Db 1830 GCCGTGGAACCCTACCATGGAAAAGTAGTGGTGCCAGAAGGACATGCAATACCCGTCCAG 1889
Qy 1864 GACTTCCAGGCATTGAGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAAC 1923
Db 1890 GACTTTCAAGCTCTGAGTGAAAGTGCCACCATTGTGTACAACGAACGAGAGTTCGTAAAC 1949
Qy 1924 AGATACCTGCACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTAT 1983
Db 1950 AGGTATCTGCACCATATTGCCACACATGGAGGAGCGCTGAACACAGATGAAGAATATTAC 2009
Qy 1984 AAGACTGTAAAGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAG 2043
Db 2010 AAAATTGTCAAGCCCAGCGAGCACGACGGCGAATACCTGTATGACATCGACAGGAAACAG 2069
Qy 2044 TGTGTTAAGCGAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCA 2103
Db 2070 TGCGTCAAGAAAGAACTAGTCACTGGGCTAGGGCTTACAGGCGAGCTGGTGGATCCTCCC 2129
Qy 2104 TTTCACGAGTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCA 2163
Db 2130 TTCCATGAATTCGCCTACGAGAGTCTGAGAACACGACCGGCCGCTCCTTACCAAGTACCA 2189
Qy 2164 ACCATCGGAGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTG 2223
Db 2190 ACCATAGGGGTGTATGGCGTGCCGGGTTCAGGCAAGTCTGGCATCATTAAAAGCGCAGTC 2249
Qy 2224 ACTAAGAAAGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGAT 2283
Db 2250 ACCAAAAAAGATCTGGTGGTGAGCGCCAAGAAAGAAAACTGCGCAGAAATAATAAGGGAC 2309
Qy 2284 GTAAGGAGGATGAGACGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAAT 2343
Db 2310 GTCAAGAAAATGAAAGGGCTGGACGTCAATGCCAGAACTGTGGACTCAGTGCTCTTGAAT 2369
Qy 2344 GGGGTTAAGCACCCCGTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGG 2403
Db 2370 GGATGCAAACACCCCGTAGAGACCCTGTATATTGACGAAGCTTTTGCTTGTCATGCAGGC 2429
Qy 2404 ACGCTGCTGGCACTGATTGCCATCGTCAAACCTAAGAAAGTGGTATTGTGCGGGGACCCA 2463
Db 2430 ACTCTCAGAGCGCTCATAGCCATCATAAGACCTAAAAAGGCAGTGCTCTGCGGGGATCCA 2489
Qy 2464 AAACAATGCGGCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGC 2523
Db 2490 AAACAGTGCGGCTTTTTCAATATGATGTGCCTGAAAGTGCATTTTAACCACGAGATTTGC 2549
Qy 2524 ACTGAAGTGTACCATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATCGTC 2583
Db 2550 ACGCAGGTCTTCCACAAAAGCATCTCTCGTCGTTGCACTAAATCTGTGACTTCGGTCGTC 2609
Qy 2584 TCCACGCTCTTCTACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATCATC 2643
Db 2610 TCAACCTTGTTTTACGACAAAAGAATGAGAACGACGAACCCGAAAGAGACTAAGATTGAG 2669
Qy 2644 ATAGATACCACAGGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGA 2703
Db 2670 ATTGACACTACTGGCAGTACTAAACCTGAGCAGGACGATCTCATTCTCACTTGTTTCAGA 2729
Qy 2704 GGATGGGTGAAACAGCTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCTGCA 2763
Db 2730 GGGTGGGTGAAGCAGTTGCAAATAGATTACAAAGGCAACGAAATAATGACGGCAGCTGCC 2789
Qy 2764 TCGCAAGGACTTACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCA 2823
Db 2790 TCTCAAGGGCTGACCCGTAAAGGCGTGTATGCCGTTCGGTACAAGGTGAATGAAAATCCC 2849
Qy 2824 CTCTACTCGCAGACTTCTGAGCACGTGAACGTGTTACTTACACGCACAGAAAAACGCATT 2883
Db 2850 CTGTACGCACCCACCTCAGAACATGTGAACGTCCTACTGACCCGCACGGAGGACCGTATC 2909
Qy 2884 GTCTGGAAGACGCTAGCTGGTGATCCCTGGATAAAGACACTTACAGCTAAATATCCCGGG 2943
Db 2910 GTGTGGAAAACACTAGCCGGCGATCCATGGATAAAAACACTGACGGCCAAGTATCCTGGG 2969
Qy 2944 GATTTCACGGCTTCATTGGACGACTGGCAGCGCGAACACGACGCCATTATGGCACGCGTT 3003
Db 2970 AATTTCACTGCCACGATAGAGGAATGGCAAGCAGAGCATGAGGCCATCATGAGGCACATC 3029
Qy 3004 CTTGATAAGCCGCAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAG 3063
Db 3030 TTGGAGAGACCGGACCCTACCGACGCATTCCAGAATAAGGCGAACGTGTGTTGGGCCAAG 3089
Qy 3064 GCTTTAGAGCCAGTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAGACG 3123
Db 3090 GCTTTGGTGCCGGTACTGAAAACTGCAGGCATAGACATGACCACTGAACAATGGAACACT 3149
Qy 3124 TTGCACCCATTCAAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTTTGC 3183
Db 3150 GTGGATTATTTCGAAACGGACAAAGCTCACTCAGCAGAGATAGTATTGAACCAACTATGC 3209
Qy 3184 ACCAGGTTCTTTGGAGTAGACCTGGACAGTGGGTTATTTTCCGCTCCTACCGTCGCACTT 3243
Db 3210 GTGAGGTTCTTTGGACTCGACCTGGACTCCGGTCTATTTTCTGCACCCACTGTTCCGTTA 3269
Qy 3244 ACTTACAGGGATCAGCACTGGGATAACTCGCCAGGGAAGAACATGTATGGGCTTAATAGA 3303
Db 3270 TCCATTAGGAATAATCACTGGGATAATTCCCCGTCGCCTAACATGTACGGGCTGAATAAA 3329
Qy 3304 GAGGTAGCAAAGGAGTTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACAGGC 3363
Db 3330 GAAGTGGTCCGTCAGCTCTCCCGCAGGTACCCACGACTGCCTCGAGCAGTTGCTACTGGA 3389
Qy 3364 AGGGTAGCTGATATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTGGTT 3423
Db 3390 AGAGTCTATGACATGAACACCGGCACACTGTGCAATTATGATCCGCGCATAAATCTAGTA 3449
Qy 3424 CCATTAAATCGCCGGTTGCCCCACTCGTTGATCGTTGACCACAAAGGACAGGGTACAACT 3483
Db 3450 CCAGTGAACAGAAGACTGCCTCATGCTTTAGTCCTCCACCATAATGAACACCCACAGAGT 3509
Qy 3484 GATCACAGCGGATTCCTATCTAAGATGAAGGGCAAATCTGTGTTGGTGATCGGCGATCCT 3543
Db 3510 GACTTTTCTTCATTCGTCAGCAAACTGAAGGGCAGAACTGTCTTGGTGGTCGGGGAGAAG 3569

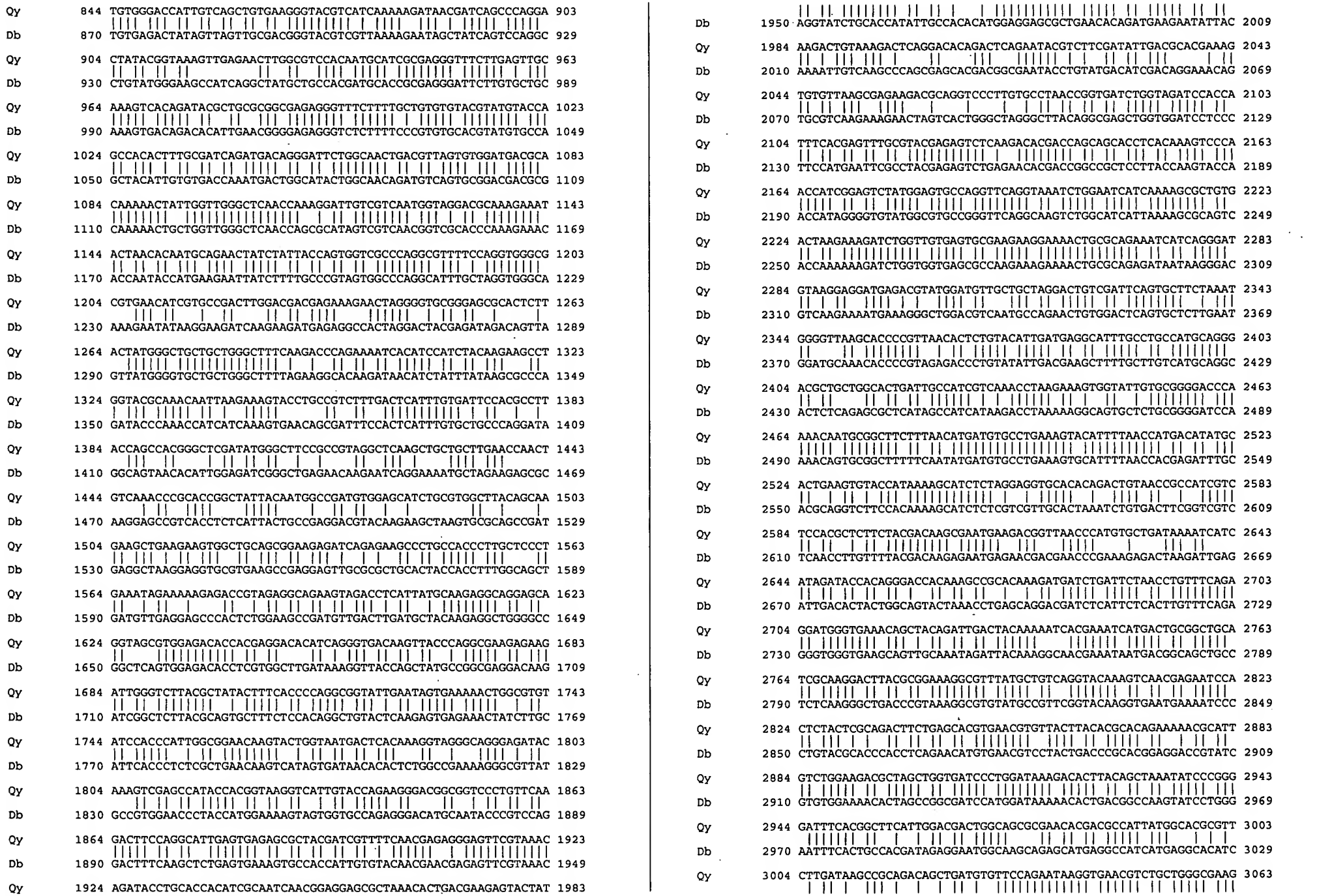
```
Qy       3544 ATCAGCATTCCAGGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATACCATCAGG 3603
Db       3570 TTGTCCGTCCCAGGCAAAACGGTTGACTGGTTGTCAGACCAGCCTGAGGCTACCTTCAGA 3629

Qy       3604 TGTGATCTCGATTTGGGAATACCTAGCCATGTCGGTAAATATGACATTATCTTTGTCAAT 3663
Db       3630 GCTCGGCTGGATTTAGGTATCCCAGGTGCCGTGCCCAAATACGACATTGTATTTATTAAT 3689

Qy       3664 GTTAGGACCCCGTACAGGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCACCAC 3723
Db       3690 GTGAGGACTCCATATAAATACCATCATTATCAGCAGTGTGAAGACCACGCCATTAAGCTT 3749

Qy       3724 AGCATGCTAACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCTATA 3783
Db       3750 AGTATGTTGACCAAGAAAGCTTGTCTGCATCTGAATCCCGGCGGAACCTGCGTCAGCATA 3809

Qy       3784 GGGTATGGGCTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCACGCTCATTT 3843
Db       3810 GGTTATGGTTACGCTGACAGGGCCAGCGAAAGCATTATTGGTGCTATAGCGCGGCAGTTC 3869

Qy       3844 AGGTTTACCCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTCGTG 3903
Db       3870 AAGTTCTCCCGGGTATGCAAACCGAAATCCTCACTTGAAGAGACAGAAGTTCTGTTTGTA 3929

Qy       3904 TTCTTCGGCAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTGCTT 3963
Db       3930 TTCATTGGGTACGATCGCAAGGCCCGTACGCACAATCCTTACAAGCTTTCATCTACCTTG 3989

Qy       3964 GACAACATCTATCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGTACAGAGTG 4023
Db       3990 ACCAACATCTATACAGGTTCCAGACTCCACGAAGCCGGATGCGCACCCTCATATCATGTG 4049

Qy       4024 ATCAGAGGTGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGCAAA 4083
Db       4050 GTGCGAGGGGACATTGCCACGGCCACCGAAGGAGTGATCATAAATGCTGCTAACAGCAAA 4109

Qy       4084 GGTCAACCAGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTTGAT 4143
Db       4110 GGACAACCTGGCGGAGGGGTGTGCGGAGCGCTGTATAAGAAATTCCCGGAAAGCTTCGAT 4169

Qy       4144 AGACAGCCAATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACC---GCTCATCATA 4200
Db       4170 TTACAGCCGATCGAAGTGGGAAAAGCGCGACTGGTCAAAGGTGCAGCTAAGCATATCATT 4229

Qy       4201 CATGCTGTAGGACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTAAGCTCGCA 4260
Db       4230 CATGCCGTAGGGCCCAACTTCAACAAAGTTTCGGAAGTTGAAGGTGACAAACAGTTGGCA 4289

Qy       4261 GCTGCCTACATGAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTA 4320
Db       4290 GAGGCCTATGAGTCCATCGCTAAAATTGTCAACGATAACAATTACAAGTCAGTAGCGATT 4349

Qy       4321 CCGCTACTGTCAACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCAT 4380
Db       4350 CCACTGTTGTCCACCGGCATCTTTTCCGGGAACAAAGATCGACTAACCCAATCATTGAAC 4409

Qy       4381 CACCTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAA 4440
Db       4410 CATTTGCTGACAGCTTTAGACACCACTGATGCAGATGTAGCCATATACTGCAGGGACAAG 4469

Qy       4441 CAATGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATTCTGGAT 4500
Db       4470 AAATGGGAAATGACTCTCAAGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGGAGATATGT 4529

Qy       4501 GATGACAAGCC----------------AGTAGACATTGACTTGGTCAGGGTCCACCCAAAC 4545
Db       4530 ATATCCGACGACTCTTCGGTGACAGAACCGGATGCAGAGCTGGTGAGGGTACATCCGAAG 4589

Qy       4546 AGCTCTTTGGCAGGCAGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTG 4605
Db       4590 AGTTCTTTGGCTGGAAGGAAGGGCTACAGCACAAGCGATGGCAAGACTTTCTCATATTTG 4649

Qy       4606 GAAGGTACACGATTCCATCAGACCGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCC 4665
Db       4650 GAAGGGACCAAATTTCACCAGGCGGCCAAGGATATAGCAGAAATTAATGCCATGTGGCCC 4709

Qy       4666 AACAAATCTGAGGCTAATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGC 4725
Db       4710 GTTGCAACGGAGGCCAATGAGCAAGTATGCATGTATATCCTTGGTGAAAGCATGAGCAGC 4769

Qy       4726 ATCCGCTCCAAATGCCCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCA 4785
Db       4770 ATTAGGTCAAAATGCCCCGTCGAAGAGTCGGAAGCCTCCACACCACCTAGCACGCTGCCT 4829

Qy       4786 TGCCTGTGTAATTACGCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAA 4845
Db       4830 TGCTTGTGCATCCATGCTATGACTCCAGAAAGAGTACAACGCCTAAAAGCCTCACGTCCA 4889

Qy       4846 GAACAGTTCGCCGTATGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAG 4905
Db       4890 GAACAAATTACTGTGTGCTCATCCTTTCCATTGCCGAAGTATAGAATCACTGGTGTGCAG 4949

Qy       4906 AAGCTACAGTGCAGCAAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCC 4965
Db       4950 AAGATCCAGTGCTCCCAGCCTATACTGTTCTCACCTAAGGTGCCTGCGTACATTCATCCA 5009

Qy       4966 AGGAAGTAC 4974
Db       5010 CGAAAGTAC 5018
```

```
RESULT 8
AF004458
LOCUS       AF004458               11420 bp    RNA     linear   VRL 12-JUL-1999
DEFINITION  Venezuelan equine encephalitis virus strain 66637, complete genome.
ACCESSION   AF004458
VERSION     AF004458.2  GI:5442458
KEYWORDS    .
SOURCE      Venezuelan equine encephalitis virus
  ORGANISM  Venezuelan equine encephalitis virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
            Alphavirus; VEEV complex.
REFERENCE   1  (bases 8366 to 9182)
  AUTHORS   Powers,A.M., Oberste,M.S., Brault,A.C., Rico-Hesse,R.,
            Schmura,S.M., Smith,J.F., Kang,W., Sweeney,W.P. and Weaver,S.C.
  TITLE     Repeated emergence of epidemic/epizootic Venezuelan equine
            encephalitis from a single genotype of enzootic subtype ID virus
  JOURNAL   J. Virol. 71 (9), 6697-6705 (1997)
  MEDLINE   97404681
   PUBMED   9261393
REFERENCE   2  (bases 1 to 11420)
  AUTHORS   Wang,E., Barrera,R., Boshell,J., Ferro,C., Freier,J.E.,
            Navarro,J.C., Salas,R., Vasquez,C. and Weaver,S.C.
  TITLE     Genetic and phenotypic changes accompanying the emergence of
            epizootic subtype IC Venezuelan equine encephalitis viruses from an
            enzootic subtype ID progenitor
  JOURNAL   J. Virol. 73 (5), 4266-4271 (1999)
  MEDLINE   99214366
   PUBMED   10196323
REFERENCE   3  (bases 8366 to 9182)
  AUTHORS   Powers,A.M., Oberste,M.S., Brault,A.C., Rico-Hesse,R.,
            Schmura,S.M., Smith,J.F., Kang,W., Sweeney,W.P. and Weaver,S.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-1997) Pathology, Univ. Texas Med. Branch,
            Galveston, TX 77555-0609, USA
REFERENCE   4  (bases 1 to 11420)
  AUTHORS   Wang,E. and Weaver,S.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-1999) Pathology, Univ. Texas Med. Branch,
            Galveston, TX 77555-0609, USA
  REMARK    Sequence update by submitter
COMMENT     On Jul 12, 1999 this sequence version replaced gi:3844617.
FEATURES             Location/Qualifiers
     source          1. .11420
                     /organism="Venezuelan equine encephalitis virus"
                     /mol_type="genomic RNA"
```

```
                /strain="66637"
                /db_xref="taxon:11036"
CDS             45. .7502
                /note="contains nonstructural proteins 1, 2, 3 and 4;
                in-frame termination codon is read through at low
                frequency"
                /codon_start=1
                /transl_except=(pos:5658. .5660,aa:OTHER)
                /product="nonstructural polyprotein"
                /protein_id="AAD43355.1"
                /db_xref="GI:5442460"
                /translation="MEKVHVDIEEDSPFLRALQRSFPQFEVEAKQVTDNDHANARAFS
                HLASKLIETEVDPSDTILDIGSAPARRMYSKHKYHCICPMRCAEDPDRLYKYATKLKK
                NFKEITDKELDKKMKELAAVMSDPDLETETMCLHDDESCRYEGQVAVYQDVYAVDGPT
                SLYHQANKGVRVAYWIGFDTTPFMFKNLAGAYPSYSTNWADETVLTARNIGLCSSDVM
                ERSRRGMSILRKKYLKPSNNVLFSVGSTIYHEKRDLLRSWHLPSVFHLRGKQNYTCRC
                ETIVSCDGYVVKRIAISPGLYGKPSGYAATMHREGFLCCKVTDTLNGERVSFPVCTYV
                PATLCDQMTGILATDVSADDAQKLLVGLNQRIVVNGRTQRNTNTMKNYLLPVVAQAFA
                RWAKEYKEDQEDERPLGLRDRQLVMGCCWAFRRHKITSIYKRPDTQTIIKVNSDFHSF
                VLPRIGSNTLEIGLRTRIRKMLEERKEPSPLITAEDVQEAKCAADEAKEVREAEELRA
                ALPPLAADVEEPTLEADVDLMLQEAGAGSVETPRGLIKVTSYAGEDKIGSYAVLSPQA
                VLKSEKLSCIHPLAEQVIVITHSGRKGRYAVEPYHGKVVVPEGHAIPVQDFQALSESA
                TIVYNEREFVNRYLHHIATHGGALNTDEEYYKIVKPSEHDGEYLYDIDRKQCVKKELV
                TGLGLTGELVDPPFHEFAYESLRTRPAAPYQVPTIGVYGVPGSGKSGIIKSAVTKKDL
                VVSAKKENCAEIIRDVKKMKGLDVNARTVDSVLLNGCKHPVETLYIDEAFACHAGTLR
                ALIAIIRPKKAVLCGDPKQCGFFNMMCLKVHFNHEICTQVFHKSISRRCTKSVTSVVS
                TLFYDKRMRTTNPKETKIEIDTTGSTKPEQDDLILTCFRGWVKQLQIDYKGNEIMTAA
                ASQGLTRKGVYAVRYKVNENPLYAPTSEHVNVLLTRTEDRIVWKTLAGDPWIKTLTAK
                YPGNFTATIEEWQAEHEAIMRHILERPDPTDAFQNKANVCWAKALVPVLKTAGIDMTT
                EQWNTVDYFETDKAHSAEIVLNQLCVRFFGLDLDSGLFSAPTVPLSIRNNHWDNSPSP
                NMYGLNKEVVRQLSRRYPRLPRAVATGRVYDMNTGTLCNYDPRINLVPVNRRLPHALV
                LHHNEHPQSDFSSFVSKLKGRTVLVVGEKLSVPGKTVDWLSDQPEATFRARLDLGIPG
                AVPKYDIVFINVRTPYKYHHYQQCEDHAIKLSMLTKKACLHLNPGGTCVSIGYGYADR
                ASESIIGAIARQFKFSRVCKPKSSLEETEVLFVFIGYDRKARTHNPYKLSSTLTNIYT
                GSRLHEAGCAPSYHVVRGDIATATEGVIINAANSKGQPGGGVCGALYKKFPESFDLQP
                IEVGKARLVKGAAKHIIHAVGPNFNKVSEVEGDKQLAEAYESIAKIVNDNNYKSVAIP
                LLSTGIFSGNKDRLTQSLNHLLTALDTTDADVAIYCRDKKWEMTLKEAVARREAVEEI
                CISDDSSVTEPDAELVRVHPKSSLAGRKGYSTSDGKTFSYLEGTKFHQAAKDIAEINA
                MWPVATEANEQVCMYILGESMSSIRSKCPVEESEASTPPSTLPCLCIHAMTPERVQRL
                KASRPEQITVCSSFPLPKYRITGVQKIQCSQPILFSPKVPAYIHPRKYLVETPPVEEN
                QSTEGTPEPAPVNVDATRTRTPEPIIIEEEEEDSISLLSDGPTHQVLQVEADIHGPPS
                VSSSSWSIPHASDFDVDSLSILDTLEGASVTSEAASAETNSYFARSMEFLARPVPAPR
                TVFRNPPHPAPRTRTPSLAPSRASSRTSLVSTPPSANRVITREELEALTPSRAPSRSA
                SRTSLVSNPPGVNRVITREEFEAFVAQQQXRFDAGAYIFSSDTGQGHLQQKSVRQTVL
                SEVVLERTELEISYAPRLDQEKEELLRKKLQLNPTPANRSRYQSRRVENMKAITARRI
                LQGLGHYLKAEGKVECYRTLHPVPLYSSSVNRAFSSPKVAVEACNAMLKENFPTVASY
                CIIPEYDAYLDMVDGASCCLDTASFCPAKLRSFPKKHSYLEPTIRSAVPSAIQNTLQN
                VLAAATKRNCNVTQMRELPVLDSAAFNVECFKKYACNNEYWETFKENPIRLTEENVVN
                YITKLKGPKAAALFAKTHNLNMLQDIPMDRFVMDLKRDVKVTPGTKHTEERPKVQVIQ
                AADPLATAYLCGIHRELVRRLNAVLLPNIHTLFDMSAEDFDAIIAEHFQPGDCVLETD
                IASFDKSEDDAMALTALMILEDLGVDAELLTLIEAAFGEISSIHLPTKTKFKFGAMMK
                SGMFLTLFVNTVINIVIASRVLRERLTGSPCAAFIGDDNIVKGVKSDKLMADRCATWL
                NMEVKIIDAVVGEKAPYFCGGFILCDTVTGTACRVADPLKRLFKLGKPLAADDEHDDD
                RRRALHEESTRWNRVGILPELCKAVESRYETVGTSIIVMAMTTLASSVKSFSYLRGAP
                ITLYG"
CDS             7538. .11305
                /note="contains capsid, E3, E2, 6K and E1 proteins"
                /codon_start=1
                /product="structural polyprotein"
                /protein_id="AAC71187.2"
                /db_xref="GI:5442459"
                /translation="MFPFQPMYPMQPMPYRNPFAAPRRPWFPRTDPFLAMQVQELTRS
                MANLTFKQRRDAPPEGPPAKKPKREAPQKQKGGGQGKKKKNQGKKKAKTGPPNPKVQS
                GNKKKTNKKPGKRQRMVMKLESDKTFPIMLEGKINGYACVVGGKLFRPMHVEGKIDND
                VLAALKTKKASKYDLEYADVPQNMRADTFKYTHEKPQGYYSWHHGAVQYENGRFTVPK
                GVGAKGDSGRPILDNQGRVVAIVLGGVNEGSRTALSVVMWNEKGVTVKYTPENCEQWS
                LVTTMCLLANVTFPCAQPPICYDRKPAETLAMLSVNVDNSGYDELLEAAVKCPGRKRR
                STEELFKEYKLTRPYMARCIRCAVGSCHSPIAIEAVKSDGHDGYVRLQTSSQYGLDSS
                GNLKGRTMRYDMHGTIEEIPLHQVSLHTSRPCHIVDGHGYFLLARCPAGDSITMEFKK
                NLVTHSCSVPYEVKFNPVGRELYTHPPEHGAEQACQVYAHDAQNRGAYVEMHLPGSEV
                DSSLVSLSGSSVTVTPPAGTSALVECECGGTKISETINTAKQFSQCTKKEQCRAYRLQ
                NDKWVYNSDKLPKAAGATLKGKLHVPFLLADGKCTVPLAPEPMITFGFRSVSLKLHPK
                NPTYLTTRQLADEPHYTHELISEPAVRNFTVTEKGWEFVWGNHPPKRFWAQETAPGNP
                HGLPHEVITHYYHRYPMSTILGLSICAAIVTVSVAASTWLFCKSRVSCLTPYRLTPNA
                RMPLCLAVLCCARTARAETIWESLDHLWNNNQQMFWIQLLIPLAALIVVTRLLKCVCC
                VVPFLVVAGAAGAGAYEHATTMPSQAGISYNTIVNRAGYAPLPISITPTKIKLIPTVN
                LEYVTCHYKTGMDSPAIKCCGSQECTPTYRPDEQCKVFTGVYPFMWGGAYCFCDTENT
                QVSKAYVMKSDDCLADHAEAYKAHTASVQAFLNITVGEHSIVTTVYVNGETPVNFNGV
                KLTAGPLSTAWTPFDRKIVQYAGEIYNYDFPEYGAGQPGAFGDIQSRTVSSSDLYANT
                NLVLQRPKAGAIHVPYTQAPSGFEQWKKDKAPSLKFTAPFGCEIYTNPIRAENCAVGS
                IPLAFDIPDALFTRVSETPTLSAAECTLNECVYSSDFGGIATVKYSASKSGKCAVHVP
                SGTATLKEAAVELTEQGSVTIHFSTANIHPEFRLQICTSYVTCKGDCHPPKDHIVTHP
                QYHAQTFTAAVSKTAWTWLTSLLGGSAVIIIIGLVLATIVAMYVLTNQKHN"
BASE COUNT      3210 a    2817 c    2906 g    2487 t
ORIGIN
```

Query Match 42.5%; Score 2123.8; DB 14; Length 11420;
Best Local Similarity 64.6%; Pred. No. 0;
Matches 3223; Conservative 0; Mismatches 1742; Indels 24; Gaps 3;

```
Qy      10 ACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTC 69
           | | | | | |||| ||||| | | |||||||||||| | || |  |||||||| |   ||
Db      30 AATTACCTACCCAAAATGGAGAAAGTTCACGTTGACATCGAGGAAGACAGCCCATTCCTC 89

Qy      70 AAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAAT 129
           |   | || || ||||  || || || |||||| | ||||| | |||||||||||| |||
Db      90 AGAGCTTTGCAACGGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCACTGATAAT 149

Qy     130 GACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTC 189
           |||||||| |||||||||||||||||||||| ||||  |||| || || || |  || ||
Db     150 GACCATGCTAATGCCAGAGCGTTTTCGCATCTGGCTTCAAAACTGATCGAAACGGAGGTG 209

Qy     190 GACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAAT 249
           ||||    | |    ||| | ||||||||||||||||||| | |      |||| ||
Db     210 GACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCCGCCCGCAGAATGTATTCTAAG 269

Qy     250 CACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGG 309
           ||    |||||||| ||||| || |||| | | || ||||| |||||||||  |   |
Db     270 CATAAGTATCATTGCATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGATTGTACAAG 329

Qy     310 TATGCAGAAAGACTTAAGAAAAGT------GACATTACCGACAAGAACATAGCCTCTAAG 363
           ||||||   | || ||||||| |      || || || |||||  | | | |   ||
Db     330 TATGCAACTAAGCTGAAGAAAAATTTCAAGGAAATAACTGACAAGGAATTGGACAAGAAA 389

Qy     364 GCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATG 423
             |   || ||    | ||||||      || |||   || |||    || ||||  |
Db     390 ATGAAGGAGCTCGCCGCCGTCATGAGCGACCCTGACCTGGAAACTGAGACTATGTGCCTC 449

Qy     424 CACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCA 483
           |||  ||||  |||| | |||   ||  || || ||||| ||||| ||||||
Db     450 CACGACGACGAGTCGTGTCGCTACGAAGGGCAAGTCGCTGTTTACCAGGATGTATACGCA 509

Qy     484 GTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGG 543
           ||  | | |||||||    |||| |||| ||     || |||||||         ||||||
Db     510 GTTGACGGACCGACAAGTCTCTATCACCAAGCCAACAAGGGAGTTAGAGTCGCCTACTGG 569

Qy     544 ATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTAC 603
           |||||||||||||| ||||||||||||  || ||| |||| ||  | || ||  | |||
Db     570 ATAGGCTTTGACACCACCCCTTTTATGTTTAAGAACTTGGCTGGAGCATATCCATCATAC 629

Qy     604 AACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCA 663
           ||  |||||||| |||| |  || ||     || |||||||| |||| | |  ||
Db     630 TCTACCAACTGGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCCTATGCAGTTCC 689

Qy     664 GATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCT 723
           ||  |  |||| |    | | ||    | || || ||||||||||||   |  ||||
Db     690 GACGTCATGGAGCGGTCACGTAGAGGGATGTCCATTCTTAGGAAGAAGTATTTGAAACCA 749

Qy     724 ACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTA 783
           | |||||  || ||||||| ||||| || || ||||||   ||  | ||    |  |
Db     750 TCCAATAATGTCCTATTCTCTGTTGGCTCGACCATCTACCACGAGAAGAGGGACTTACTG 809

Qy     784 CGTAGCTGGCATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGA 843
           | |||||||| || ||    || || |||||   || |||   || | |||| || |
Db     810 AGGAGCTGGCACCTGCCGTCTGTATTTCACTTACGTGGCAAGCAAAATTACACATGTCGG 869
```

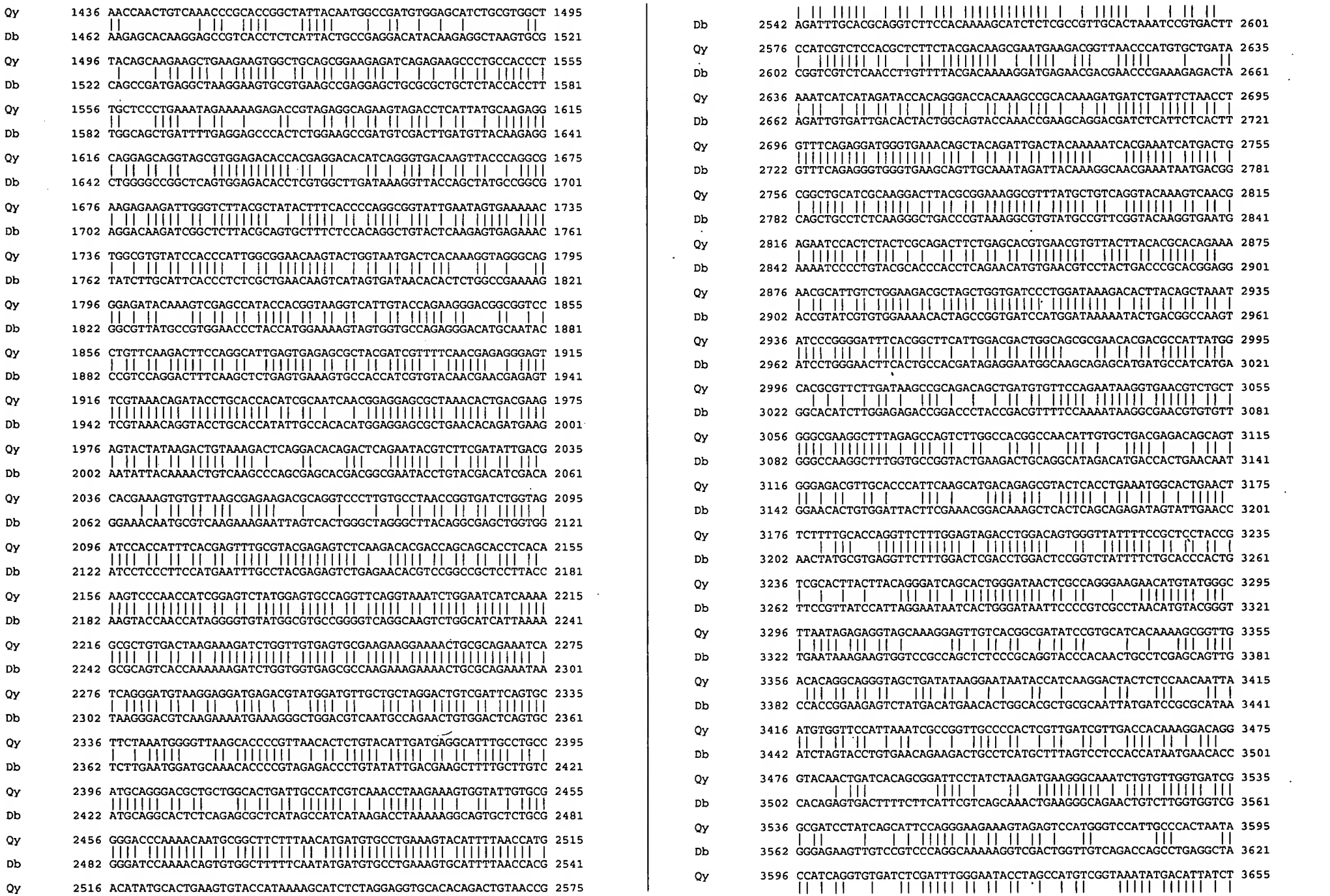


```
Db       3030 TTGGAGAGACCGGACCCTACCGACGCATTCCAGAATAAGGCGAACGTGTGTTGGGCCAAG 3089
Qy       3064 GCTTTAGAGCCAGTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAGACG 3123
Db       3090 GCTTTGGTGCCGGTACTGAAAACTGCAGGCATAGACATGACCACTGAACAATGGAACACT 3149
Qy       3124 TTGCACCCATTCAAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTTTGC 3183
Db       3150 GTGGATTATTTCGAAACGGACAAAGCTCACTCAGCAGAGATAGTATTGAACCAACTATGC 3209
Qy       3184 ACCAGGTTCTTTGGAGTAGACCTGGACAGTGGGTTATTTTCCGCTCCTACCGTCGCACTT 3243
Db       3210 GTGAGGTTCTTTGGACTCGACCTGGACTCCGGTCTATTTTCTGCACCCACTGTTCCGTTA 3269
Qy       3244 ACTTACAGGGATCAGCACTGGGATAACTCGCCAGGGAAGAACATGTATGGGCTTAATAGA 3303
Db       3270 TCCATTAGGAATAATCACTGGGATAATTCCCCGTCGCCTAACATGTACGGGCTGAATAAA 3329
Qy       3304 GAGGTAGCAAAGGAGTTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACAGGC 3363
Db       3330 GAAGTGGTCCGTCAGCTCTCCCGCAGGTACCCACGACTGCCTCGAGCAGTTGCTACTGGA 3389
Qy       3364 AGGGTAGCTGATATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTGGTT 3423
Db       3390 AGAGTCTATGACATGAACACCGGCACACTGTGCAATTATGATCCGCGCATAAATCTAGTA 3449
Qy       3424 CCATTAAATCGCCGGTTGCCCCACTCGTTGATCGTTGACCACAAAGGACAGGGTACAACT 3483
Db       3450 CCAGTGAACAGAAGACTGCCTCATGCTTTAGTCCTCCACCATAATGAACACCCACAGAGT 3509
Qy       3484 GATCACAGCGGATTCCTATCTAAGATGAAGGGCAAATCTGTGTTGGTGATCGGCGATCCT 3543
Db       3510 GACTTTTCTTCGTTCGTCAGCAAACTGAAGGGCAGAACTGTCTTGGTGGTCGGGGAGAAG 3569
Qy       3544 ATCAGCATTCCAGGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATACCATCAGG 3603
Db       3570 TTGTCCGTCCCAGGCAAAACGGTTGACTGGTTGTCAGACCAGCCTGAGGCTACCTTCAGA 3629
Qy       3604 TGTGATCTCGATTTGGGAATACCTAGCCATGTCGGTAAATATGACATTATCTTTGTCAAT 3663
Db       3630 GCTCGGCTGGATTTAGGTATCCCAGGTGCCGTGCCCAAATACGACATTGTATTTATTAAT 3689
Qy       3664 GTTAGGACCCCGTACAGGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCACCAC 3723
Db       3690 GTGAGGACTCCATATAAATACCATCATTATCAGCAGTGTGAAGACCACGCCATTAAGCTT 3749
Qy       3724 AGCATGCTAACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCTATA 3783
Db       3750 AGTATGTTGACCAAGAAAGCTTGTCTGCATCTGAATCCCGGCGGAACCTGCGTCAGCATA 3809
Qy       3784 GGGTATGGGCTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCACGCTCATTT 3843
Db       3810 GGTTATGGTTACGCTGACAGGGCCAGCGAAAGCATTATTGGTGCTATAGCGCGGCAGTTC 3869
Qy       3844 AGGTTTACCCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTCGTG 3903
Db       3870 AAGTTCTCCCGGGTATGCAAACCGAAATCCTCACTTGAAGAGACAGAAGTTCTGTTTGTA 3929
Qy       3904 TTCTTCGGCAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTGCTT 3963
Db       3930 TTCATTGGGTACGATCGCAAGGCCCGTACGCACAATCCTTACAAGCTTTCATCTACCTTG 3989
Qy       3964 GACAACATCTATCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGTACAGAGTG 4023
Db       3990 ACCAACATCTATACAGGTTCCAGACTCCACGAAGCCGGATGCGCACCCTCATATCATGTG 4049
Qy       4024 ATCAGAGGTGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGCAAA 4083
Db       4050 GTGCGAGGGGACATTGCCACGGCCACCGAAGGAGTGATCATAAATGCTGCTAACAGCAAA 4109
Qy       4084 GGTCAACCAGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTTGAT 4143
Db       4110 GGACAACCTGGCGGAGGGGTGTGCGGAGCGCTGTATAAGAAATTCCCGGAAAGCTTCGAT 4169

Qy       4144 AGACAGCCAATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACC---GCTCATCATA 4200
Db       4170 TTACAGCCGATCGAAGTGGGAAAAGCGCGACTGGTCAAAGGTGCAGCTAAGCATATCATT 4229
Qy       4201 CATGCTGTAGGACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTAAGCTCGCA 4260
Db       4230 CATGCCGTAGGGCCCAACTTCAACAAAGTTTCGGAAGTTGAAGGTGACAAACAGTTGGCA 4289
Qy       4261 GCTGCCTACATGAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTA 4320
Db       4290 GAGGCCTATGAGTCCATCGCTAAAATTGTCAACGATAACAATTACAAGTCAGTAGCGATT 4349
Qy       4321 CCGCTACTGTCAACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCAT 4380
Db       4350 CCACTGTTGTCCACCGGCATCTTTTCCGGGAACAAAGATCGACTAACCCAATCATTGAAC 4409
Qy       4381 CACCTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAA 4440
Db       4410 CATTTGCTGACAGCTTTAGACACCACTGATGCAGATGTAGCCATATACTGCAGGGACAAG 4469
Qy       4441 CAATGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATTCTGGAT 4500
Db       4470 AAATGGGAAATGACTCTCAAGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGGAGATATGT 4529
Qy       4501 GATGACAAGCC----------------AGTAGACATTGACTTGGTCAGGGTCCACCCAAAC 4545
Db       4530 ATATCCGACGACTCTTCGGTGACAGAACCGGATGCAGAGCTGGTGAGGGTACATCCGAAG 4589
Qy       4546 AGCTCTTTGGCAGGCAGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTG 4605
Db       4590 AGTTCTTTGGCTGGAAGGAAGGGCTACAGCACAAGCGATGGCAAGACTTTCTCATATTTG 4649
Qy       4606 GAAGGTACACGATTCCATCAGACCGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCC 4665
Db       4650 GAAGGGACCAAATTTCACCAGGCGGCCAAGGATATAGCAGAAATTAATGCCATGTGGCCC 4709
Qy       4666 AACAAATCTGAGGCTAATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGC 4725
Db       4710 GTTGCAACGGAGGCCAATGAGCAAGTATGCATGTATATCCTTGGTGAAAGCATGAGCAGC 4769
Qy       4726 ATCCGCTCCAAATGCCCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCA 4785
Db       4770 ATTAGGTCAAAATGCCCCGTCGAAGAGTCGGAAGCCTCCACACCACCTAGCACGCTGCCT 4829
Qy       4786 TGCCTGTGTAATTACGCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAA 4845
Db       4830 TGCTTGTGCATCCATGCTATGACTCCAGAAAGAGTACAACGCCTAAAAGCCTCACGTCCA 4889
Qy       4846 GAACAGTTCGCCGTATGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAG 4905
Db       4890 GAACAAATTACTGTGTGCTCATCCTTTCCATTGCCGAAGTATAGAATCACTGGTGTGCAG 4949
Qy       4906 AAGCTACAGTGCAGCAAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCC 4965
Db       4950 AAGATCCAGTGCTCCCAGCCTATACTGTTCTCACCTAAGGTGCCTGCGTACATTCATCCA 5009
Qy       4966 AGGAAGTAC 4974
Db       5010 CGAAAGTAC 5018
```

```
RESULT 9
EEVCOMGEN
LOCUS       EEVCOMGEN              11444 bp ss-RNA     linear   VRL 01-JUN-2001
DEFINITION  Venezuelan equine encephalitis virus, complete genome.
ACCESSION   L04653 L00931
VERSION     L04653.1  GI:290609
KEYWORDS    .
SOURCE      Venezuelan equine encephalitis virus
  ORGANISM  Venezuelan equine encephalitis virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
            Alphavirus; VEEV complex.
```

```
REFERENCE   1  (bases 1 to 11444)
  AUTHORS   Kinney,R.M., Tsuchiya,K.R., Sneider,J.M. and Trent,D.W.
  TITLE     Genetic evidence that epizootic Venezuelan equine encephalitis
            (VEE) viruses may have evolved from enzootic VEE subtype I-D virus
  JOURNAL   Virology 191 (2), 569-580 (1992)
  MEDLINE   93079859
   PUBMED   1448915
REFERENCE   2  (bases 1 to 11444)
  AUTHORS   Kinney,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-1993) Molecular Virology, Centers for Disease
            Control and Prevention, PO Box 2087, Fort Collins, CO 80522, USA
FEATURES             Location/Qualifiers
     source          1. .11444
                     /organism="Venezuelan equine encephalitis virus"
                     /mol_type="genomic RNA"
                     /strain="P676"
                     /isolate="mosquitos/Venezuela/1963"
                     /db_xref="taxon:11036"
                     /tissue_lib="pUC18 and M13 of Kinney et al."
     mRNA            1. .11444
                     /function="genomic mRNA of VEE virus as determined from
                     strain P676"
     gene            1. .7526
                     /gene="NS"
     5'UTR           1. .44
                     /gene="NS"
     CDS             45. .7526
                     /gene="NS"
                     /note="possible incorporation of arginine, cysteine or
                     tryptophan at read through of UGA codon"
                     /codon_start=1
                     /transl_except=(pos:5682. .5684,aa:Arg)
                     /product="putative nonstructural polyprotein precursor"
                     /protein_id="AAC19318.1"
                     /db_xref="GI:3201593"
                     /translation="MEKVHVDIEEDSPFLRALQRSFPQFEVEAKQVTDNDHANARAFS
                     HLASKLIETEVDPSDTILDIGSAPARRMYSKHKYHCICPMRCAEDPDRLYKYATKLKK
                     NCKEITDKELDKKMKELAAVMSDPDLETETMCLHDDESCRYEGQVAVYQDVYAVDGPT
                     SLYHQANKGVRVAYWIGFDTTPFMFKNLAGAYPSYSTNWADETVLTARNIGLCSSDVM
                     ERSRRGMSILRKKYLKPSNNVLFSVGSTIYHEKRDLLRSWHLPSVFHLRGKQNYTCRC
                     ETIVSCDGYVVKRIAISPGLYGKPSGYAATMHREGFLCCKVTDTLNGERVSFPVCTYV
                     PATLCDQMTGILATDVSADDAQKLLVGLNQRIVVNGRTQRNTNTMKNYLLPVVAQAFA
                     RWAKEYKEDQEDERPLGLRDRQLVMGCCWAFRRHKITSIYKRPDTQTIIKVNSDFHSF
                     VLPRIGSNTLEIGLRTRIRKMLEEHKEPSPLITAEDIQEAKCAADEAKEVREAEELRA
                     ALPPLAADFEEPTLEADVDLMLQEAGAGSVETPRGLIKVTSYAGEDKIGSYAVLSPQA
                     VLKSEKLSCIHPLAEQVIVITHSGRKGRYAVEPYHGKVVVPEGHAIPVQDFQALSESA
                     TIVYNEREFVNRYLHHIATHGGALNTDEEYYKTVKPSEHDGEYLYDIDRKQCVKKELV
                     TGLGLTGELVDPPFHEFAYESLRTRPAAPYQVPTIGVYGVPGSGKSGIIKSAVTKKDL
                     VVSAKKENCAEIIRDVKKMKGLDVNARTVDSVLLNGCKHPVETLYIDEAFACHAGTLR
                     ALIAIIRPKKAVLCGDPKQCGFFNMMCLKVHFNHEICTQVFHKSISRRCTKSVTSVVS
                     TLFYDKRMRTTNPKETKIVIDTTGSTKPKQDDLILTCFRGWVKQLQIDYKGNEIMTAA
                     ASQGLTRKGVYAVRYKVNENPLYAPTSEHVNVLLTRTEDRIVWKTLAGDPWIKILTAK
                     YPGNFTATIEEWQAEHDAIMRHILERPDPTDVFQNKANVCWAKALVPVLKTAGIDMTT
                     EQWNTVDYFETDKAHSAEIVLNQLCVRFFGLDLDSGLFSAPTVPLSIRNNHWDNSPSP
                     NMYGLNKEVVRQLSRRYPQLPRAVATGRVYDMNTGTLRNYDPRINLVPVNRRLPHALV
                     LHHNEHPQSDFSSFVSKLKGRTVLVVGEKLSVPGKKVDWLSDQPEATFRARLDLGIPG
                     DVPKYDIVFINVRTPYKYHHYQQCEDHAIKLSMLTKKACLHLNPGGTCVSIGYGYADR
                     ASESIIGAIARQFKFSRVCKPKSSHEETEVLFVFIGYDRKARTHNPYKLSSTLTNIYT
                     GSRLHEAGCAPSYHVVRGDIATATEGVIINAANSKGQPGGGVCGALYKKFPESFDLQP
                     IEVGKARLVKGAAKHIIHAVGPNFNKVSEVEGDKQLAEAYESIAKIVNDNNYKSVAIP
                     LLSTGIFSGNKDRLTQSLNHLLTALDTTDADVAIYCRDKKWEMTLKEAVARREAVEEI
                     CISDDSSVTEPDAELVRVHPKSSLAGRKGYSTSDGKTFSYLEGTKFHQAAKDIAEINA
                     MWPVATEANEQVCMYILGESMSSIRSKCPVEESEASTPPSTLPCLCIHAMTPERVQRL
                     KASRPEQITVCSSFPLPKYRITGVQKIQCSQPILFSPKVPAYIHPRKYLVETPPVEET
                     PESPAENQSTEGTPEQPALVNVDATRTRMPEPIIIEEEEEDSISLLSDGPTHQVLQVE
                     ADIHGSPSVSSSSWSIPHASDFDVDSLSILDTLDGASVTSGAVSAETNSYFARSMEFR
                     ARPVPAPRTVFRNPPHPAPRTRTPPLAHSRASSRTSLVSTPPGVNRVITREELEALTP
                     SRAPSRSASRTSLVSNPPGVNRVITREEFEAFVAQQQRRFDAGAYIFSSDTGQGHLQQ
                     KSVRQTVLSEVVLERTELEISYAPRLDQEKEELLRKKLQLNPTPANRSRYQSRRVENM
                     KAITARRILQGLGHYLKAEGKVECYRTLHPVPLYSSSVNRAFSSPKVAVEACNAMLKE
                     NFPTVASYCIIPEYDAYLDMVDGASCCLDTASFCPAKLRSFPKKHSYLEPTIRSAVPS
                     AIQNTLQNVLAAATKRNCNVTQMRELPVLDSAAFNVECFKKYACNNEYWETFKENPIR
                     LTEENVVNYITKLKGPKAAALFAKTHNLNMLQDIPMDRFVMDLKRDVKVTPGTKHTEE
                     RPKVQVIQAADPLATADLCGIHRELVRRLNAVLLPNIHTLFDMSAEDFDAIIAEHFQP
                     GDCVLETDIASFDKSEDDAMALTALMILEDLGVDAELLTLIEAAFGEISSIHLPTKTK
                     FKFGAMMKSGMFLTLFVNTVINIVIASRVLRERLTGSPCAAFIGDDNIVKGVKSDKLM
                     ADRCATWLNMEVKIIDAVVGEKAPYFCGGFILCDSVTGTACRVADPLKRLFKLGKPLA
                     VDDEHDDDRRRALHEESTRWNRVGILPELCKAVESRYETVGTSIIVMAMTTLASSVKS
                     FSYLRGAPITLYG"
     mat_peptide     45. .1649
                     /gene="NS"
                     /product="putative nonstructural protein nsP1"
                     /function="minus strand RNA synthesis; methyltransferase;
                     guanyltransferase"
     mat_peptide     1650. .4031
                     /gene="NS"
                     /product="putative nonstructural protein nsP2"
                     /function="RNA helicase; nonstructural proteinase;
                     necessary for subgenomic 26S mRNA synthesis"
                     /function="replication"
     mat_peptide     4032. .5702
                     /gene="NS"
                     /product="putative nonstructural protein nsP3"
                     /function="replication"
     mat_peptide     5703. .7520
                     /gene="NS"
                     /product="putative nonstructural protein nsP4"
                     /function="RNA polymerase"
                     /function="replication"
     CDS             45. .5684
                     /gene="NS"
                     /codon_start=1
                     /product="nonstructural polyprotein precursor"
                     /protein_id="AAC19320.1"
                     /db_xref="GI:3201594"
                     /translation="MEKVHVDIEEDSPFLRALQRSFPQFEVEAKQVTDNDHANARAFS
                     HLASKLIETEVDPSDTILDIGSAPARRMYSKHKYHCICPMRCAEDPDRLYKYATKLKK
                     NCKEITDKELDKKMKELAAVMSDPDLETETMCLHDDESCRYEGQVAVYQDVYAVDGPT
                     SLYHQANKGVRVAYWIGFDTTPFMFKNLAGAYPSYSTNWADETVLTARNIGLCSSDVM
                     ERSRRGMSILRKKYLKPSNNVLFSVGSTIYHEKRDLLRSWHLPSVFHLRGKQNYTCRC
                     ETIVSCDGYVVKRIAISPGLYGKPSGYAATMHREGFLCCKVTDTLNGERVSFPVCTYV
                     PATLCDQMTGILATDVSADDAQKLLVGLNQRIVVNGRTQRNTNTMKNYLLPVVAQAFA
                     RWAKEYKEDQEDERPLGLRDRQLVMGCCWAFRRHKITSIYKRPDTQTIIKVNSDFHSF
                     VLPRIGSNTLEIGLRTRIRKMLEEHKEPSPLITAEDIQEAKCAADEAKEVREAEELRA
                     ALPPLAADFEEPTLEADVDLMLQEAGAGSVETPRGLIKVTSYAGEDKIGSYAVLSPQA
                     VLKSEKLSCIHPLAEQVIVITHSGRKGRYAVEPYHGKVVVPEGHAIPVQDFQALSESA
                     TIVYNEREFVNRYLHHIATHGGALNTDEEYYKTVKPSEHDGEYLYDIDRKQCVKKELV
                     TGLGLTGELVDPPFHEFAYESLRTRPAAPYQVPTIGVYGVPGSGKSGIIKSAVTKKDL
                     VVSAKKENCAEIIRDVKKMKGLDVNARTVDSVLLNGCKHPVETLYIDEAFACHAGTLR
                     ALIAIIRPKKAVLCGDPKQCGFFNMMCLKVHFNHEICTQVFHKSISRRCTKSVTSVVS
                     TLFYDKRMRTTNPKETKIVIDTTGSTKPKQDDLILTCFRGWVKQLQIDYKGNEIMTAA
                     ASQGLTRKGVYAVRYKVNENPLYAPTSEHVNVLLTRTEDRIVWKTLAGDPWIKILTAK
                     YPGNFTATIEEWQAEHDAIMRHILERPDPTDVFQNKANVCWAKALVPVLKTAGIDMTT
                     EQWNTVDYFETDKAHSAEIVLNQLCVRFFGLDLDSGLFSAPTVPLSIRNNHWDNSPSP
                     NMYGLNKEVVRQLSRRYPQLPRAVATGRVYDMNTGTLRNYDPRINLVPVNRRLPHALV
                     LHHNEHPQSDFSSFVSKLKGRTVLVVGEKLSVPGKKVDWLSDQPEATFRARLDLGIPG
                     DVPKYDIVFINVRTPYKYHHYQQCEDHAIKLSMLTKKACLHLNPGGTCVSIGYGYADR
                     ASESIIGAIARQFKFSRVCKPKSSHEETEVLFVFIGYDRKARTHNPYKLSSTLTNIYT
                     GSRLHEAGCAPSYHVVRGDIATATEGVIINAANSKGQPGGGVCGALYKKFPESFDLQP
                     IEVGKARLVKGAAKHIIHAVGPNFNKVSEVEGDKQLAEAYESIAKIVNDNNYKSVAIP
                     LLSTGIFSGNKDRLTQSLNHLLTALDTTDADVAIYCRDKKWEMTLKEAVARREAVEEI
                     CISDDSSVTEPDAELVRVHPKSSLAGRKGYSTSDGKTFSYLEGTKFHQAAKDIAEINA
                     MWPVATEANEQVCMYILGESMSSIRSKCPVEESEASTPPSTLPCLCIHAMTPERVQRL
                     KASRPEQITVCSSFPLPKYRITGVQKIQCSQPILFSPKVPAYIHPRKYLVETPPVEET
                     PESPAENQSTEGTPEQPALVNVDATRTRMPEPIIIEEEEEDSISLLSDGPTHQVLQVE
                     ADIHGSPSVSSSSWSIPHASDFDVDSLSILDTLDGASVTSGAVSAETNSYFARSMEFR
                     ARPVPAPRTVFRNPPHPAPRTRTPPLAHSRASSRTSLVSTPPGVNRVITREELEALTP
                     SRAPSRSASRTSLVSNPPGVNRVITREEFEAFVAQQQ"
     misc_feature    7524. .7561
                     /note="putative"
                     /function="noncoding segment between nonstructural and
                     structural genes"
     gene            7532. .11444
                     /gene="26S mRNA"
     mRNA            7532. .11444
```

```
                     /gene="26S mRNA"
                     /product="26S mRNA region"
                     /note="The structural proteins of the virus are translated
                     as apolyprotein precursor molecule from an intracellular,
                     26SmRNA species that is identical to the 3'-one-third
                     portion of the viral genomic mRNA; putative"
                     /function="intracellular, subgenomic viral mRNA species"
     CDS             7562. .11329
                     /gene="26S mRNA"
                     /codon_start=1
                     /product="structural polyprotein precursor"
                     /protein_id="AAC19319.1"
                     /db_xref="GI:290611"
                     /translation="MFPFQPMYPMQPMPYRNPFAAPRRPWFPRTDPFLAMQVQELTRS
                     MANLTFKQRRDAPPEGPPAKKPKREAPQKQKGGGQGKKKKNQGKKKAKTGPPNPKAQS
                     GNKKKPNKKPGKRQRMVMKLESDKTFPIMLEGKINGYACVVGGKLFRPMHVEGKIDND
                     VLAALKTKKASKYDLEYADVPQNMRADTFKYTHEKPQGYYSWHHGAVQYENGRFTVPK
                     GVGAKGDSGRPILDNQGRVVAIVLGGVNEGSRTALSVVMWNEKGVTVKYTPENCEQWS
                     LVTTMCLLANVTFPCAEPPICYDRKPAETLAMLSVNVDNPGYDELLEAAVKCPGRKRR
                     STEELFKEYKLTRPYMARCIRCAVGSCHSPIAIEAVKSDGHDGYVRLQTSSQYGLDSS
                     GNLKGRTMRYDMHGTIEEIPLHQVSLHTSRPCHIVDGHGYFLLARCPAGDSITMEFKK
                     GSVTHSCSVPYEVKFNPVGRELYTHPPEHGAEQACQVYAHDAQNRGAYVEMHLPGSEV
                     DSSLISLSGSSVTVTPPVGTSALVKCKCGGTKISETINKAKQFSQCTKKEQCRAYRLQ
                     NDKWVYNSDKLPKAAGATLKGKLHVPFLLADGKCTVPLAPEPMITFGFRSVSLKLHPK
                     NPTYLTTRQLADEPHYTHELISEPAVRNFTVTEKGWEFVWGNHPPKRFWAQETAPGNP
                     HGLPHEVITHYYHRYPMSTILGLSICAAIVTVSVAASTWLFCKSRVSCLTPYRLTPNA
                     RMPLCLAVLCCARTARAETTWESLDHLWNNNQQMFWIQLLIPLAALIVVTRLLKCVCC
                     VVPFLVVAGAAGAGAYEHATTMPSQAGISYNTIVNRAGYAPLPISITPTKIKLIPTVN
                     LEYVTCHYKTGMDSPAIKCCGSQECTPTNRPDEQCKVFTGVYPFMWGGAYCFCDTENT
                     QVSKAYVMKSDDCLADHAEAYKAHTASVQAFLNITVGEHSIVTTVYVNGETPVNFNGV
                     KLTAGPLSTAWTPFDRKIVQYAGEIYNYDFPEYGAGQPGAFGDIQSRTVSSSDLYANT
                     NLVLQRPKAGAIHVPYTQAPSGFEQWKKDKAPSLKFTAPFGCEIYTNPIRAENCAVGS
                     IPLAFDIPDALFTRVSETPTLSAAECTLNECVYSSDFGGIATVKYSASKSGKCAVHVP
                     SGTATLKEAAVELTEQGSATIHFSTANIHPEFRLQICTSYVTCKGDCHPPKDHIVTHP
                     QYHAQTFTAAVSKTAWTWLTSLLGGSAVIIIIGLVLATIVAMYVLTNQKHN"
     mat_peptide     7562. .8386
                     /gene="26S mRNA"
                     /product="core protein"
                     /function="encloses the genomic mRNA molecule"
                     /standard_name="capsid protein"
                     /note="putative"
     mat_peptide     8387. .8563
                     /gene="26S mRNA"
                     /product="viral cleavage product"
                     /standard_name="E3 protein"

  Query Match             42.4%;  Score 2122.2;  DB 14;  Length 11444;
  Best Local Similarity   64.5%;  Pred. No. 0;
  Matches 3225;  Conservative    0;  Mismatches 1748;  Indels   24;  Gaps    3;

Qy          2 CCCTACAAACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCC 61
              |||  |   | | | |  |||| |||||  | | ||||||||||| | ||  |  |||||||
Db         22 CCCAAACCAATTACCTACCCAAAATGGAGAAAGTTCACGTTGACATCGAGGAAGACAGCC 81

Qy         62 CGTATGTCAAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCA 121
              | |   |||   | ||||| ||||  || || || |||||| | ||||| | ||||||||
Db         82 CATTCCTCAGAGCTTTACAACGGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCA 141

Qy        122 CTGACAATGACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGA 181
              |||| |||||||||| |||||||||||||||||||||| ||||  |||| || || || |
Db        142 CTGATAATGACCATGCTAATGCCAGAGCGTTTTCGCATCTGGCTTCAAAACTGATCGAAA 201

Qy        182 GCGAAGTCGACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCAC 241
              || || ||||    | |    ||| | ||||||||||||||||||| | |
Db        202 CGGAGGTGGACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCCGCCCGCAGAATGT 261

Qy        242 ATTCCAATCACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGAC 301
              |||| || ||     |||||||| ||||| || |||| | | || ||||| |||||||||
Db        262 ATTCTAAGCATAAGTATCATTGCATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGAT 321

Qy        302 TACAACGGTATGCAGAAAGACTTAAGAAAAGT------GACATTACCGACAAGAACATAG 355
              |  |   |||||||    |  || |||||||  |       || || ||| |||||| |  | |
Db        322 TGTACAAGTATGCAACTAAGCTGAAGAAAAATTGCAAGGAAATAACTGACAAGGAATTGG 381

Qy        356 CCTCTAAGGCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTC 415
              |   ||  |  || ||    | ||||||      || |||   || |||    ||
Db        382 ACAAGAAAATGAAGGAGCTCGCCGCCGTCATGAGCGACCCTGACCTGGAAACTGAGACTA 441

Qy        416 TGTGTATGCACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATG 475
              ||||  | |||   || |   | ||| | |||   ||    || || || ||||| ||||
Db        442 TGTGCCTCCACGACGATGAGTCATGTCGCTACGAGGGGCAAGTCGCTGTTTACCAGGATG 501

Qy        476 TGTACGCAGTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAA 535
              | ||||||||  | | |||||||    |||| ||||| ||    || |||||||||
Db        502 TATACGCAGTTGACGGACCGACAAGTCTCTATCACCAAGCCAACAAGGGAGTTAGAGTCG 561

Qy        536 TTTACTGGATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACC 595
                |||||||||||||||||||| ||||||||||||  || ||| |||| ||  | || |
Db        562 CCTACTGGATAGGCTTTGACACCACCCCTTTTATGTTTAAGAACTTGGCTGGAGCATATC 621

Qy        596 CTACTTACAACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCG 655
              |  | |||   || |||||||| ||||| |  || ||    || |||||||| |||||
Db        622 CATCATACTCTACCAACTGGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCCTAT 681

Qy        656 GTAACTCAGATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGC 715
              | | ||| ||  |   |||| |    | | ||    | || || ||||||||||||
Db        682 GCAGCTCCGACGTCATGGAGCGGTCACGTAGAGGGATGTCCATTCTTAGGAAGAAGTATT 741

Qy        716 TCCAACCTACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGAT 775
              |  |||| | |||||  || ||||||| ||||| || || ||||||   ||  | ||
Db        742 TGAAACCATCCAATAATGTCCTATTCTCTGTTGGCTCGACCATCTACCACGAGAAGAGGG 801

Qy        776 CACTGTTACGTAGCTGGCATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCA 835
              |  |  | |||||||| || ||    || || |||||    || |||   || | ||
Db        802 ACTTACTGAGGAGCTGGCACCTGCCGTCTGTATTTCACTTACGTGGCAAGCAAAATTACA 861

Qy        836 CAGGTAGATGTGGGACCATTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCA 895
              || || | |||| ||| || || || || || ||||||||| | |||| ||| | ||||
Db        862 CATGTCGGTGTGAGACTATAGTTAGTTGCGACGGGTACGTCGTTAAAAGAATAGCTATCA 921

Qy        896 GCCCAGGACTATACGGTAAAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCT 955
              | ||||| || || || ||   |    ||  ||  | || ||||| ||||||||  ||||
Db        922 GTCCAGGCCTGTATGGGAAGCCTTCAGGCTATGCTGCTACGATGCACCGCGAGGGATTCT 981

Qy        956 TGAGTTGCAAAGTCACAGATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGT 1015
              || | |||||||| |||||  ||  ||| ||||||||  |||||| | |||||  ||||
Db        982 TGTGCTGCAAAGTGACAGACACATTGAACGGGGAGAGGGTCTCTTTTCCCGTGTGCACGT 1041

Qy       1016 ATGTACCAGCCACACTTTGCGATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGG 1075
              |||| ||||| ||| | || || ||||| || || |||||||  || || |||| || ||
Db       1042 ATGTGCCAGCTACATTGTGTGACCAAATGACTGGCATACTGGCAACAGATGTCAGTGCGG 1101

Qy       1076 ATGACGCACAAAAACTATTGGTTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGC 1135
              | ||||| ||||||||  |||||||||||||||  | || ||||||||  ||| | || |
Db       1102 ACGACGCGCAAAAACTGCTGGTTGGGCTCAACCAGCGCATAGTCGTCAACGGTCGCACCC 1161

Qy       1136 AAAGAAATACTAACACAATGCAGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCA 1195
              |||||| || || || ||| |||| ||||| || || || || ||||||||  ||| | |
Db       1162 AAAGAAACACCAATACCATGAAGAATTATCTTTTGCCCGTAGTGGCCCAGGCATTTGCTA 1221

Qy       1196 GGTGGGCGCGTGAACATCGTGCCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGC 1255
              |||||||     ||| ||    |  ||    || || |||||      ||||||  | || ||
Db       1222 GGTGGGCAAAGGAATATAAGGAAGATCAAGAAGATGAGAGGCCACTAGGACTACGAGATA 1281

Qy       1256 GCACTCTTACTATGGGCTGCTGCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACA 1315
              |    |  ||||| |||||||||||||  |  || || || ||||| || | ||
Db       1282 GACAGTTAGTCATGGGGTGCTGCTGGGCTTTTAGAAGGCACAAGATAACATCTATTTATA 1341

Qy       1316 AGAAGCCTGGTACGCAAACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTC 1375
              ||   || | ||| |||||  || |  |||||       ||  ||  |||||| ||| | |
Db       1342 AGCGCCCAGATACCCAAACCATCATCAAAGTGAACAGCGATTTCCACTCATTCGTGCTGC 1401

Qy       1376 CACGCCTTACCAGCCACGGGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTG 1435
              |  |  |  ||| ||  || || || ||  |  |    ||  |||  |    |||| |
Db       1402 CCAGGATAGGCAGTAACACACTGGAGATCGGGCTGAGAACGAGAATCAGGAAAATGCTAG 1461
```

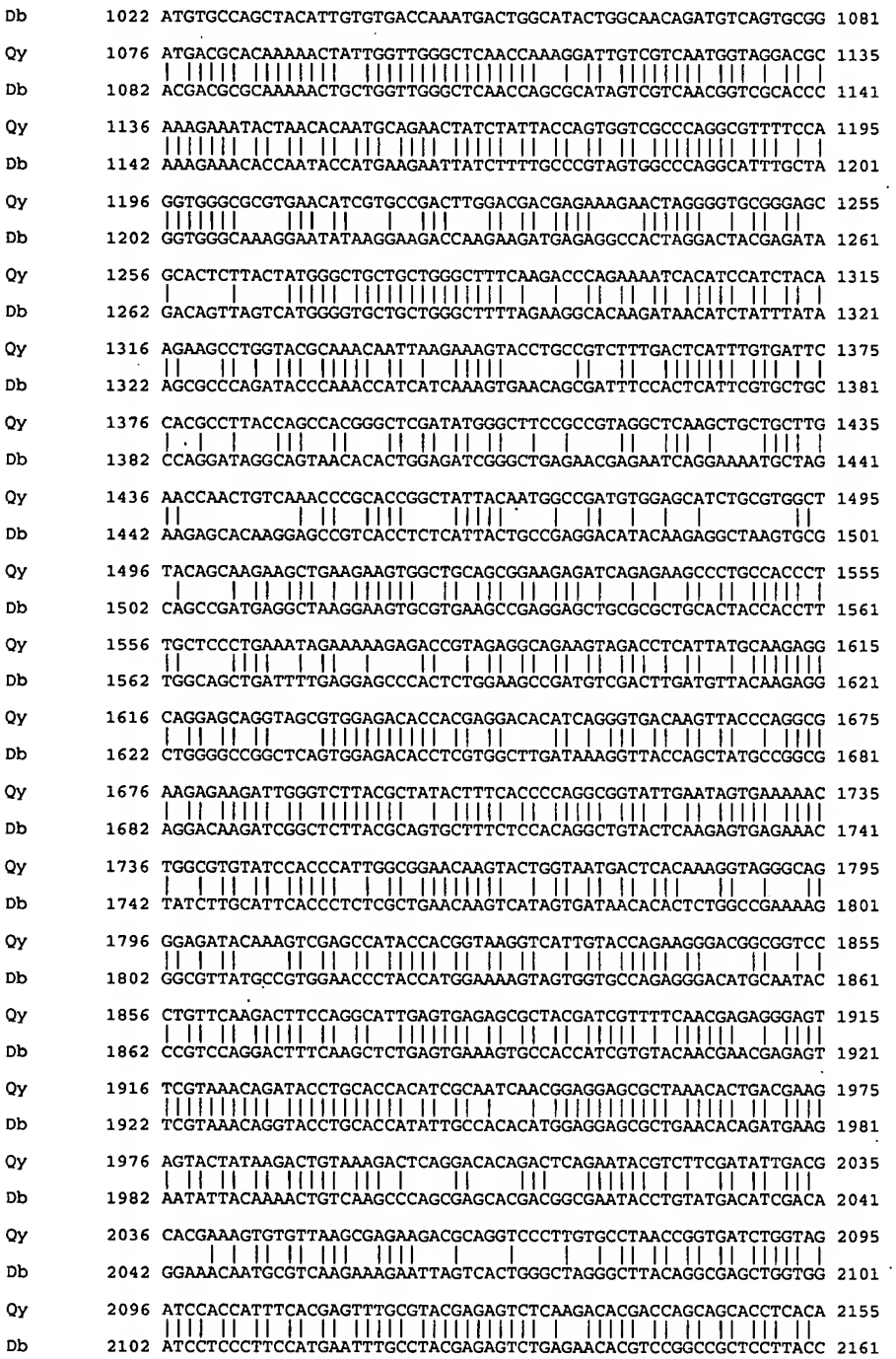


```
Db       3622 CCTTTAGAGCTCGGCTGGATTTAGGTATCCCAGGTGACGTGCCCAAATACGACATTGTAT 3681

Qy       3656 TTGTCAATGTTAGGACCCCGTACAGGAACCATCACTACCAACAGTGCGAGGATCACGCTA 3715

Db       3682 TTATTAACGTGAGGACTCCATATAAATACCATCATTATCAGCAGTGTGAAGACCACGCCA 3741

Qy       3716 TCCACCACAGCATGCTAACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTG 3775

Db       3742 TTAAGCTTAGTATGTTGACCAAGAAAGCTTGTCTGCATTTGAATCCCGGCGGAACCTGCG 3801

Qy       3776 TGGCTATAGGGTATGGGCTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCAC 3835

Db       3802 TCAGCATAGGTTATGGTTACGCTGACAGGGCCAGCGAGAGCATCATTGGTGCTATAGCGC 3861

Qy       3836 GCTCATTTAGGTTTACCCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTC 3895

Db       3862 GGCAGTTCAAGTTCTCCCGGGTATGCAAACCGAAATCCTCACATGAAGAGACAGAAGTAC 3921

Qy       3896 TCTTCGTGTTCTTCGGCAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTG 3955

Db       3922 TGTTTGTATTCATTGGGTACGATCGCAAGGCCCGTACGCACAATCCTTACAAGCTTTCAT 3981

Qy       3956 TAGTGCTTGACAACATCTATCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGT 4015

Db       3982 CTACCTTGACCAACATCTATACAGGTTCCAGACTCCACGAAGCCGGATGCGCACCCTCAT 4041

Qy       4016 ACAGAGTGATCAGAGGTGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTA 4075

Db       4042 ATCATGTGGTGCGAGGGGATATTGCCACGGCCACCGAAGGAGTGATCATAAATGCTGCTA 4101

Qy       4076 ATAGCAAAGGTCAACCAGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTG 4135

Db       4102 ACAGCAAAGGACAACCTGGCGGAGGGGTGTGCGGAGCGCTGTATAAGAAATTCCCGGAAA 4161

Qy       4136 CTTTTGATAGACAGCCAATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACC---GC 4192

Db       4162 GCTTCGATTTACAGCCGATCGAAGTAGGAAAAGCGCGACTGGTCAAAGGTGCAGCTAAAC 4221

Qy       4193 TCATCATACATGCTGTAGGACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTA 4252

Db       4222 ATATCATTCATGCCGTAGGACCAAACTTCAACAAAGTTTCGGAAGTTGAAGGGGACAAAC 4281

Qy       4253 AGCTCGCAGCTGCCTACATGAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAA 4312

Db       4282 AGTTGGCAGAGGCTTATGAGTCCATCGCTAAAATTGTCAACGATAACAATTACAAGTCAG 4341

Qy       4313 TATCAGTACCGCTACTGTCAACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAAT 4372

Db       4342 TAGCGATTCCACTGTTGTCCACCGGCATCTTTTCCGGGAACAAAGATCGACTAACCCAAT 4401

Qy       4373 CATTGCATCACCTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCT 4432

Db       4402 CATTGAACCATTTGCTGACAGCTTTAGACACCACTGATGCAGATGTAGCCATATACTGCA 4461

Qy       4433 TGGATAAACAATGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAA 4492

Db       4462 GGGACAAGAAATGGGAAATGACTCTCAAGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGG 4521

Qy       4493 TTCTGGA---------------TGATGACAAGCCAGTAGACATTGACTTGGTCAGGGTCC 4537

Db       4522 AGATATGCATATCAGACGACTCTTCGGTGACAGAACCGGATGCAGAGCTGGTGAGGGTAC 4581

Qy       4538 ACCCAAACAGCTCTTTGGCAGGCAGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTATT 4597

Db       4582 ATCCGAAGAGTTCTTTGGCTGGAAGGAAGGGCTACAGCACAAGTGATGGCAAGACTTTCT 4641

Qy       4598 CATACCTGGAAGGTACACGATTCCATCAGACCGCCAAGGACATTGCCGAAATCCATGCAA 4657

Db       4642 CATATTTGGAAGGGACCAAATTTCACCAGGCGGCCAAGGATATAGCAGAAATTAATGCCA 4701

Qy       4658 TGTGGCCCAACAAATCTGAGGCTAATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTA 4717

Db       4702 TGTGGCCAGTTGCAACGGAGGCCAATGAGCAAGTATGCATGTATATCCTCGGTGAAAGCA 4761

Qy       4718 TGTCCAGCATCCGCTCCAAATGCCCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACA 4777

Db       4762 TGAGCAGCATTAGGTCGAAATGCCCCGTCGAGGAGTCGGAAGCCTCCACACCACCTAGCA 4821

Qy       4778 CACTTCCATGCCTGTGTAATTACGCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTG 4837

Db       4822 CGCTGCCTTGCTTGTGCATCCATGCTATGACTCCAGAAAGAGTACAACGCCTAAAAGCCT 4881

Qy       4838 CGAAGAAAGAACAGTTCGCCGTATGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAG 4897

Db       4882 CACGTCCAGAACAAATTACTGTGTGCTCATCCTTTCCATTGCCGAAGTATAGAATCACTG 4941

Qy       4898 GCGTGCAGAAGCTACAGTGCAGCAAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTG 4957

Db       4942 GTGTGCAGAAGATCCAGTGCTCCCAGCCTATACTGTTCTCACCGAAGGTGCCTGCGTACA 5001

Qy       4958 TACACCCCAGGAAGTAC 4974

Db       5002 TTCATCCACGGAAGTAC 5018
```

RESULT 10
VEU55347
LOCUS VEU55347 11423 bp RNA linear VRL 18-JUN-2002
DEFINITION Venezuelan equine encephalitis virus strain 6119 nonstructural polyprotein and structural polyprotein genes, complete cds.
ACCESSION U55347
VERSION U55347.2 GI:20800451
KEYWORDS .
SOURCE Venezuelan equine encephalitis virus
ORGANISM Venezuelan equine encephalitis virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae; Alphavirus; VEEV complex.
REFERENCE 1 (bases 1 to 11423)
AUTHORS Weaver,S.C., Salas,R., Rico-Hesse,R., Ludwig,G.V., Oberste,M.S., Boshell,J. and Tesh,R.B.
TITLE Re-emergence of epidemic Venezuelan equine encephalomyelitis in South America. VEE Study Group
JOURNAL Lancet 348 (9025), 436-440 (1996)
MEDLINE 96332169
PUBMED 8709783
REFERENCE 2 (bases 1 to 11423)
AUTHORS Weaver,S.C.
TITLE Direct Submission
JOURNAL Submitted (18-APR-1996) Scott C. Weaver, Pathology, Univ. TX Med. Branch, 301 University Ave., Galveston, TX 77555-0609, USA
REFERENCE 3 (bases 1 to 11423)
AUTHORS Weaver,S.C.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-2001) Scott C. Weaver, Pathology, Univ. TX Med. Branch, 301 University Ave., Galveston, TX 77555-0609, USA
REMARK Sequence update by submitter
COMMENT On May 15, 2002 this sequence version replaced gi:2335117.
FEATURES Location/Qualifiers
source 1. .11423
/organism="Venezuelan equine encephalitis virus"
/mol_type="genomic RNA"
/strain="6119"
/db_xref="taxon:11036"
/note="antigenic IC"
CDS 25. .7506
/note="contains low efficiency read through at stop codon; encodes nsP1, nsP2, nsP3 and nsP4 proteins"
/codon_start=1
/transl_except=(pos:5662. .5664,aa:OTHER)
/product="nonstructural polyprotein"
/protein_id="AAM28637.1"
/db_xref="GI:20800453"
/translation="MEKVHVDIEEDSPFLRALQRSFPQFEVEAKQVTDNDHANARAFS
HLASKLIETEVDPSDTILDIGSAPARRMYSKHKYHCICPMRCAEDPDRLYKYATKLKK
NCKEITDKELDKKMKELAAVMSDPDLETETMCLHDDESCRYEGQVAVYQDVYAVDGPT
SLYHQANKGVRVAYWIGFDTTPFMFKNLAGAYPSYSTNWADETVLTARNIGLCSSDVM

```
                     ERSRRGMSILRKKYLKPSNNVLFSVGSTIYHEKRDLLRSWHLPSVFHLRGKQNYTCRC
                     ETIVSCDGYVVKRIAISPGLYGKPSGYAATMHREGFLCCKVTDTLNGERVSFPVCTYV
                     PATLCDQMTGILATDVSADDAQKLLVGLNQRIVVNGRTQRNTNTMKNYLLPVVAQAFA
                     RWAKEYKEDQEDERPLGLRDRQLVMGCCWAFRRHKITSIYKRPDTQTIIKVNSDFHSF
                     VLPRIGSNTLEIGLRTRIRKMLEEHKEPSPLITAEDIQEAKCAADEAKEVREAEELRA
                     ALPPLAADFEEPTLEADVDLMLQEAGAGSVETPRGLIKVTSYAGEDKIGSYAVLSPQA
                     VLKSEKLSCIHPLAEQVIVITHSGRKGRYAVEPYHGKVVVPEGHAIPVQDFQALSESA
                     TIVYNEREFVNRYLHHIATHGGALNTDEEYYKTVKPSEHDGEYLYDIDRKQCVKKELV
                     TGLGLTGELVDPPFHEFAYESLRTRPAAPYQVPTIGVYGVPGSGKSGIIKSAVTKKDL
                     VVSAKKENCAEIIRDVKKMKGLDVNARTVDSVLLNGCKHPVETLYIDEAFACHAGTLR
                     ALIAIIRPKKAVLCGDPKQCGFFNMMCLKVHFNHEICTQVFHKSISRRCTKSVTSVVS
                     TLFYDKRMRTTNPKETKIEIDTTGSTKPKQDDLILTCFRGWVKQLQIDYKGNEIMTAA
                     ASQGLTRKGVYAVRYKVNENPLYAPTSEHVNVLLTRTEDRIVWKTLAGDPWIKTLTAK
                     YPGNFTATIEEWQAEHDAIMRHILERPDPTDVFQNKANVCWAKALVPVLKTAGIDMTT
                     EQWNTVDYFETDKAHSAEIVLNQLCVRFFGLDLDSGLFSAPTVPLSIRNNHWDNSPSP
                     NMYGLNKEVVRQLSRRYPQLPRAVATGRVYDMNTGTLRNYDPRINLVPVNRRLPHALV
                     LHHNEHPQSDFSSFVSKLKGRTVLVVGEKLSVPGKKVDWLSDQPEATFRARLDLGIPG
                     DVPKYDIVFINVRTPYKYHHYQQCEDHAIKLSMLTKKACLHLNPGGTCVSIGYGYADR
                     ASESIIGAIARQFKFSRVCKPKSSHEETEVLFVFIGYDRKARTHNPYKLSSTLTNIYT
                     GSRLHEAGCAPSYHVVRGDIATATEGVIINAANSKGQPGGGVCGALYKKFPESFDLQP
                     IEVGKARLVKGVTKHIIHAVGPNFNKVSEVEGDKQLAEAYESIAKIVNDNNYKSVAIP
                     LLSTGIFSGNKDRLTQSLNHLLTALDTTDADVAIYCRDKKWEMTLKEAVARREAVEEI
                     CISDDSSVTEPDAELVRVHPKSSLAGRKGYSTSDGKTFSYLEGTKFHQAAKDIAEINA
                     MWPVATEANEQVCMYILGESMSSIRSKCPVEESEASTPPSTLPCLCIHAMTPERVQRL
                     KASRPEQITVCSSFPLPKYRITGVQKIQCSQPILFSPKVPAYIHPRKYLVETPPVEET
                     PESPAENQSTEGTPEQPALVNVDATRTRMPEPIIIEEEEEDSISLLSDGPTHQVLQVE
                     ADIHGSPSVSSSSWSIPHASDFDVDSLSILDTLDGASVTSEAVSAETNSYFARSMEFR
                     ARPVPAPRTVFRNPPHPAPRTRTPPLAHSRASSRTSLVSTPPGVNRVITREELEALTP
                     SRAPSRSASRTSLVSNPPGVNRVITREEFEAFVAQQQXRFDAGAYIFSSDTGQGHLQQ
                     KSVRQTVLSEVVLERTELEISYAPRLDQEKEELLRKKLQLNPTPANRSRYQSRRVENM
                     KAITARRILQQLGHYLKAEGKVECYRTLHPVPLYSSSVNRAFSSPKVAVEACNAMLKE
                     NFPTVASYCIIPEYDAYLDMVDGASCCLDTASFCPAKLRSFPKKHSYLEPTIRSAVPS
                     AIQNTLQNVLAAATKRNCNVTQMRELPVLDSAAFNVECFKKYACNNEYWETFKENPIR
                     LTEENVVNYITKLKGPKAAALFAKTHNLNMLQDIPMDRFVMDLKRDVKVTPGTKHTEE
                     RPKVQVIQAADPLATAYLCGIHRELVRRLNAVLLPNIHTLFDMSAEDFDAIIAEHFQP
                     GDCVLETDIASFDKSEDDAMALTALMILEDLGVDAELLTLIEAAFGEISSIHLPTKTK
                     FKFGAMMKSGMFLTLFVNTVINIVIASRVLRERLTGSPCAAFIGDDNIVKGVKSDKLM
                     ADRCATWLNMEVKIIDAVVGEKAPYFCGGFILCDSVTGTACRVADPLKRLFKLGKPLA
                     VDDEHDDDRRRALHEESTRWNRVGILPELCKAVESRYETVGTSIIVMAMTTLASSVKS
                     FSYLRGAPITLYG"
     CDS             7542. .11309
                     /note="encodes capsid protein and E2 and E1 envelope
                     glycoproteins"
                     /codon_start=1
                     /product="structural polyprotein"
                     /protein_id="AAC58464.2"
                     /db_xref="GI:20800452"
                     /translation="MFPFQPMYPMQPMPYRNPFAAPRRPWFPRTDPFLAMQVQELTRS
                     MANLTFKQRRDAPPEGPPAKKPKREAPQKQKGGGQGKKKNQGKKKAKTGPPNPKAQS
                     GNKKKPNKKPGKRQRMVMKLESDKTFPIMLEGKINGYACVVGGKLFRPMHVEGKIDND
                     VLAALKTKKASKYDLEYADVPQNMRADTFKYTHEKPQGYYSWHHGAVQYENGRFTVPK
                     GVGAKGDSGRPILDNQGRVVAIVLGGVNEGSRTALSVVMWNEKGVTVKYTPENCEQWS
                     LVTTMCLLANVTFPCAEPPICYDRKPAETLAMLSVNVDNPGYDELLEAAVKCPGKRR
                     STEELFKEYKLTRPYMARCIRCAVGSCHSPIAIEAVKSDGHDGYVRLQTSSQYGLDSS
                     GNLKGRTMRYDMHGTIEEIPLHQVSLHTSRPCHIVDGHGYFLLARCPAGDSITMEFKK
                     GSVTHSCSVPYEVKFNPVGRELYTHPPEHGAEQACQVYAHDAQNRGAYVEMHLPGSEV
                     DSSLISLSGSSVTVTPPVGTSALVECKCGGTKISETINKAKQFSQCTKKEQCRAYRLQ
                     NDKWVYNSDKLPKAAGATLKGKLHVPFLLADGKCTVPLAPEPMITFGFRSVSLKLHPK
                     NPTYLTTRQLADEPHYTHELISEPAVRNFTVTEKGWEFVWGNHPPKRFWAQETAPGNP
                     HGLPHEVITHYYHRYPMSTILGLSICAAIVTVSVAASTWLFCKSRVSCLTPYRLTPNA
                     RMPLCLAVLCCARTARAETTWESLDHLWNNNQQMFWIQLLIPLAALIVVTRLLKCVCC
                     VVPFLVVAGAAGAGAYEHATTMPSQAGISYNTIVNRAGYAPLPISITPTKIKLIPTVN
                     LEYVTCHYKTGMDSPAIKCCGSQECTPTNRPDEQCKVFTGVYPFMWGGAYCFCDTENT
                     QVSKAYVMKSDDCLADHAEAYKAHTASVQAFLNITVGEHSIVTTVYVNGETPVNFNGV
                     KLTAGPLSTAWTPFDRKIVQYAGEIYNYDFPEYGAGQPGAFGDIQSRTVSSSDLYANT
                     NLVLQRPKAGAIHVPYTQAPSGFEQWKKDKAPSLKFTAPFGCEIYTNPIRAENCAVGS
                     IPLAFDIPDALFTRVSETPTLSAAECTLNECVYSSDFGGIATVKYSASKSGKCAVHVP
                     SGTATLKEAAVELTEQGSATIHFSTANIHPEFRLQICTSYVTCKGDCHPPKDHIVTHP
                     QYHAQTFTAAVSKTAWTWLTSLLGGSAVIIIIGLVLATIVAMYVLTNQKHN"
BASE COUNT     3227 a   2819 c   2902 g   2475 t
ORIGIN

  Query Match           42.4%;  Score 2120.6;  DB 14;  Length 11423;
```

```
  Best Local Similarity   64.5%;  Pred. No. 0;
  Matches 3224;  Conservative    0;  Mismatches 1749;  Indels   24;  Gaps    3;

Qy          2 CCCTACAAACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCC 61
              |||  |   | | | | | |||| ||||| | | |||||||||||| | || |  |||||||
Db          2 CCCAAACCAATTACCTACCCAAAATGGAGAAAGTTCACGTTGACATCGAGGAAGACAGCC 61

Qy         62 CGTATGTCAAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCA 121
              | |   |||   | ||||| ||||  || || || |||||| | ||||| | ||||||||
Db         62 CATTCCTCAGAGCTTTACAACGGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCA 121

Qy        122 CTGACAATGACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGA 181
              |||| ||||||||||| ||||||||||||||||||||| ||||  |||| || || || |
Db        122 CTGATAATGACCATGCTAATGCCAGAGCGTTTTCGCATCTGGCTTCAAAACTGATCGAAA 181

Qy        182 GCGAAGTCGACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCAC 241
              || || |||     | |   ||| | |||||||||||||||||||| | |
Db        182 CGGAGGTGGACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCCGCCCGCAGAATGT 241

Qy        242 ATTCCAATCACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGAC 301
              |||| || ||    |||||||| ||||| || |||| | | || ||||| |||||||||
Db        242 ATTCTAAGCATAAGTATCATTGCATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGAT 301

Qy        302 TACAACGGTATGCAGAAAGACTTAAGAAAAGT------GACATTACCGACAAGAACATAG 355
              | |  |||||||  |  || |||||||| |      || || || |||||| | | |
Db        302 TGTACAAGTATGCAACTAAGCTGAAGAAAAATTGCAAGGAAATAACTGACAAGGAATTGG 361

Qy        356 CCTCTAAGGCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTC 415
              |  ||  |  || || ||    ||||||     ||||| || |||    || |||    ||
Db        362 ACAAGAAAATGAAGGAGCTCGCCGCCGTCATGAGCGACCCAGACCTGGAAACTGAGACTA 421

Qy        416 TGTGTATGCACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATG 475
              ||||  | |||   || |   | ||| | |||   ||    || || || ||||| ||||
Db        422 TGTGCCTCCACGACGATGAGTCATGTCGCTACGAGGGGCAAGTCGCTGTTTACCAGGATG 481

Qy        476 TGTACGCAGTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAA 535
              | |||||||  | | |||||||    |||| ||||| ||    || ||||||||
Db        482 TATACGCAGTTGACGGACCGACAAGTCTCTATCACCAAGCCAACAAGGGAGTTAGAGTCG 541

Qy        536 TTTACTGGATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACC 595
              ||||||||||||||||||||| |||||||||||| || ||| |||| ||  | || |
Db        542 CCTACTGGATAGGCTTTGACACCACCCCTTTTATGTTTAAGAACTTGGCTGGAGCATATC 601

Qy        596 CTACTTACAACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCG 655
              |  | |||   || |||||||| ||||  |  || ||    || |||||||| |||||
Db        602 CATCATACTCTACCAACTGGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCCTAT 661

Qy        656 GTAACTCAGATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGC 715
              | | ||| ||  |  |||| |   | | ||   | || || ||||||||||||
Db        662 GCAGCTCCGACGTCATGGAGCGGTCACGTAGAGGGATGTCCATTCTTAGGAAGAAGTATT 721

Qy        716 TCCAACCTACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGAT 775
              |  |||| | ||||| || |||||| ||||| || || |||||  ||  | ||
Db        722 TGAAACCATCCAATAATGTCCTATTCTCTGTTGGCTCGACCATCTACCACGAGAAGAGGG 781

Qy        776 CACTGTTACGTAGCTGGCATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCA 835
              |  |  | |||||||| || ||    || || |||||    || |||   || | ||
Db        782 ACTTACTGAGGAGCTGGCACCTGCCGTCTGTATTTCACTTACGTGGCAAGCAAAATTACA 841

Qy        836 CAGGTAGATGTGGGACCATTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCA 895
              || || | |||| ||| || || || || || ||||||||| | ||||  ||| | ||||
Db        842 CATGTCGGTGTGAGACTATAGTTAGTTGCGACGGGTACGTCGTTAAAAGAATAGCTATCA 901

Qy        896 GCCCAGGACTATACGGTAAAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCT 955
              | ||||| || || || ||    ||  ||  | || |||| ||||||| ||||
Db        902 GTCCAGGCCTGTATGGGAAGCCTTCAGGCTATGCTGCTACGATGCACCGCGAGGGATTCT 961

Qy        956 TGAGTTGCAAAGTCACAGATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGT 1015
              || | |||||||| ||||| ||  ||  ||| |||||||| |||||| | ||||| ||||
Db        962 TGTGCTGCAAAGTGACAGACACATTGAACGGGGAGAGGGTCTCTTTTCCCGTGTGCACGT 1021

Qy       1016 ATGTACCAGCCACACTTTGCGATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGG 1075
              |||| ||||| ||| | || || || ||||| || || |||||||| || || |||| ||
```

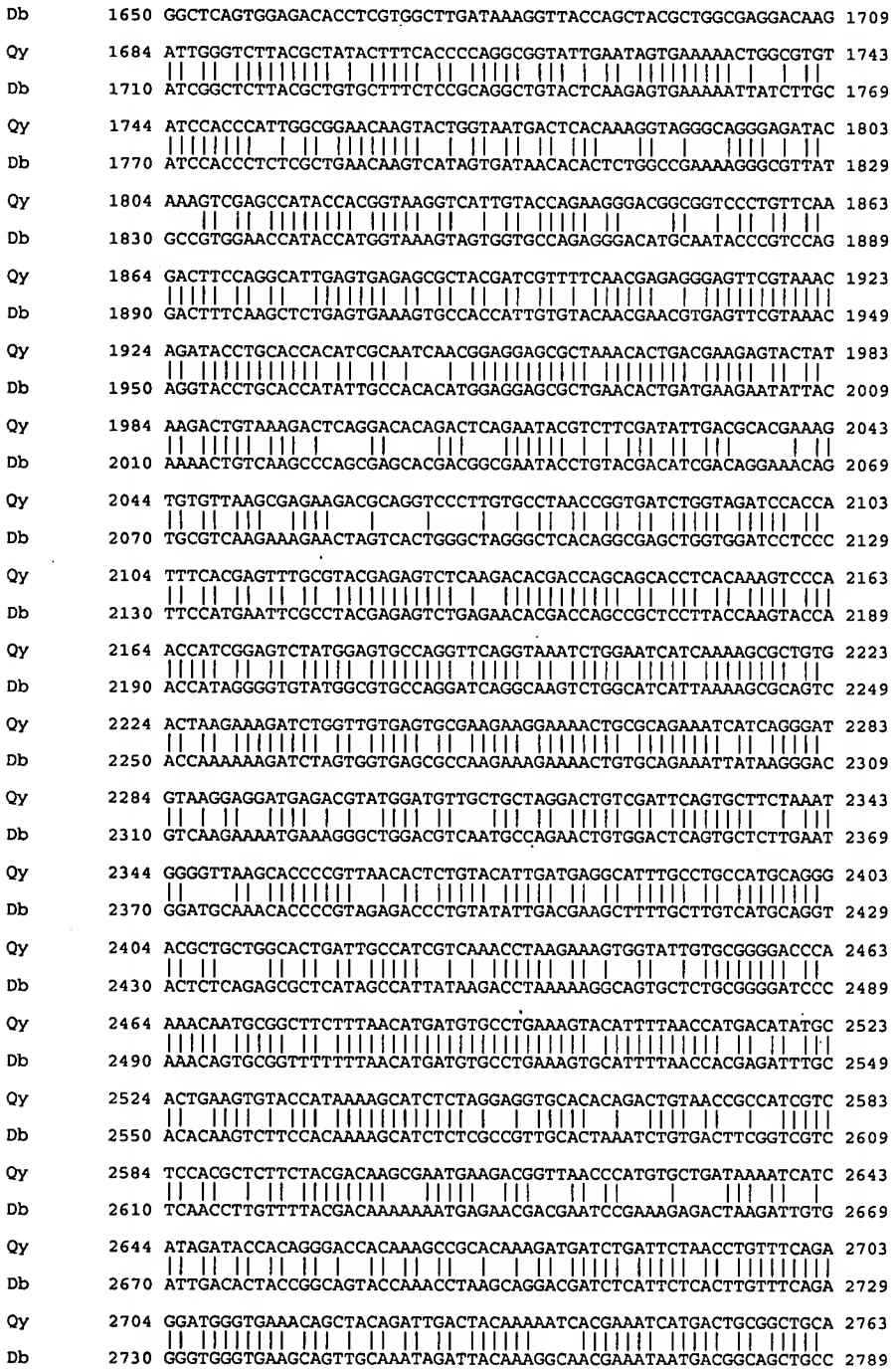

Db 1022 ATGTGCCAGCTACATTGTGTGACCAAATGACTGGCATACTGGCAACAGATGTCAGTGCGG 1081
Qy 1076 ATGACGCACAAAAACTATTGGTTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGC 1135
Db 1082 ACGACGCGCAAAAACTGCTGGTTGGGCTCAACCAGCGCATAGTCGTCAACGGTCGCACCC 1141
Qy 1136 AAAGAAATACTAACACAATGCAGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCA 1195
Db 1142 AAAGAAACACCAATACCATGAAGAATTATCTTTTGCCCGTAGTGGCCCAGGCATTTGCTA 1201
Qy 1196 GGTGGGCGCGTGAACATCGTGCCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGC 1255
Db 1202 GGTGGGCAAAGGAATATAAGGAAGACCAAGAAGATGAGAGGCCACTAGGACTACGAGATA 1261
Qy 1256 GCACTCTTACTATGGGCTGCTGCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACA 1315
Db 1262 GACAGTTAGTCATGGGGTGCTGCTGGGCTTTTAGAAGGCACAAGATAACATCTATTTATA 1321
Qy 1316 AGAAGCCTGGTACGCAAACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTC 1375
Db 1322 AGCGCCCAGATACCCAAACCATCATCAAAGTGAACAGCGATTTCCACTCATTCGTGCTGC 1381
Qy 1376 CACGCCTTACCAGCCACGGGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTG 1435
Db 1382 CCAGGATAGGCAGTAACACACTGGAGATCGGGCTGAGAACGAGAATCAGGAAAATGCTAG 1441
Qy 1436 AACCAACTGTCAAACCCGCACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCT 1495
Db 1442 AAGAGCACAAGGAGCCGTCACCTCTCATTACTGCCGAGGACATACAAGAGGCTAAGTGCG 1501
Qy 1496 TACAGCAAGAAGCTGAAGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCT 1555
Db 1502 CAGCCGATGAGGCTAAGGAAGTGCGTGAAGCCGAGGAGCTGCGCGCTGCACTACCACCTT 1561
Qy 1556 TGCTCCCTGAAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGG 1615
Db 1562 TGGCAGCTGATTTTGAGGAGCCCACTCTGGAAGCCGATGTCGACTTGATGTTACAAGAGG 1621
Qy 1616 CAGGAGCAGGTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCG 1675
Db 1622 CTGGGGCCGGCTCAGTGGAGACACCTCGTGGCTTGATAAAGGTTACCAGCTATGCCGGCG 1681
Qy 1676 AAGAGAAGATTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAAC 1735
Db 1682 AGGACAAGATCGGCTCTTACGCAGTGCTTTCTCCACAGGCTGTACTCAAGAGTGAGAAAC 1741
Qy 1736 TGGCGTGTATCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAG 1795
Db 1742 TATCTTGCATTCACCCTCTCGCTGAACAAGTCATAGTGATAACACACTCTGGCCGAAAAG 1801
Qy 1796 GGAGATACAAAGTCGAGCCATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCC 1855
Db 1802 GGCGTTATGCCGTGGAACCCTACCATGGAAAAGTAGTGGTGCCAGAGGGACATGCAATAC 1861
Qy 1856 CTGTTCAAGACTTCCAGGCATTGAGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGT 1915
Db 1862 CCGTCCAGGACTTTCAAGCTCTGAGTGAAAGTGCCACCATCGTGTACAACGAACGAGAGT 1921
Qy 1916 TCGTAAACAGATACCTGCACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAG 1975
Db 1922 TCGTAAACAGGTACCTGCACCATATTGCCACACATGGAGGAGCGCTGAACACAGATGAAG 1981
Qy 1976 AGTACTATAAGACTGTAAAGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGACG 2035
Db 1982 AATATTACAAAACTGTCAAGCCCAGCGAGCACGACGGCGAATACCTGTATGACATCGACA 2041
Qy 2036 CACGAAAGTGTGTTAAGCGAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAG 2095
Db 2042 GGAAACAATGCGTCAAGAAAGAATTAGTCACTGGGCTAGGGCTTACAGGCGAGCTGGTGG 2101
Qy 2096 ATCCACCATTTCACGAGTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACA 2155
Db 2102 ATCCTCCCTTCCATGAATTTGCCTACGAGAGTCTGAGAACACGTCCGGCCGCTCCTTACC 2161
Qy 2156 AAGTCCCAACCATCGGAGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAA 2215
Db 2162 AAGTACCAACCATAGGGGTGTATGGCGTGCCGGGGTCAGGCAAGTCTGGCATCATTAAAA 2221
Qy 2216 GCGCTGTGACTAAGAAAGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCA 2275
Db 2222 GCGCAGTCACCAAAAAAGATCTGGTGGTGAGCGCCAAGAAAGAAAACTGCGCAGAAATAA 2281
Qy 2276 TCAGGGATGTAAGGAGGATGAGACGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGC 2335
Db 2282 TAAGGGACGTCAAGAAAATGAAAGGGCTGGACGTCAATGCCAGAACTGTGGACTCAGTGC 2341
Qy 2336 TTCTAAATGGGGTTAAGCACCCCGTTAACACTCTGTACATTGATGAGGCATTTGCCTGCC 2395
Db 2342 TCTTGAATGGATGCAAACACCCCGTAGAGACCCTGTATATTGACGAAGCTTTTGCTTGTC 2401
Qy 2396 ATGCAGGGACGCTGCTGGCACTGATTGCCATCGTCAAACCTAAGAAAGTGGTATTGTGCG 2455
Db 2402 ATGCAGGCACTCTCAGAGCGCTCATAGCCATCATAAGACCTAAAAAGGCAGTGCTCTGCG 2461
Qy 2456 GGGACCCAAAACAATGCGGCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATG 2515
Db 2462 GGGATCCAAAACAGTGTGGCTTTTTCAATATGATGTGCCTGAAAGTGCATTTTAACCACG 2521
Qy 2516 ACATATGCACTGAAGTGTACCATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCG 2575
Db 2522 AGATTTGCACGCAGGTCTTCCACAAAAGCATCTCTCGTCGTTGCACTAAATCCGTGACTT 2581
Qy 2576 CCATCGTCTCCACGCTCTTCTACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATA 2635
Db 2582 CGGTCGTCTCAACCTTGTTTTACGACAAAAGGATGAGAACGACGAACCCGAAAGAGACTA 2641
Qy 2636 AAATCATCATAGATACCACAGGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCT 2695
Db 2642 AGATTGAGATTGACACTACTGGCAGTACCAAACCGAAGCAGGACGATCTCATTCTCACTT 2701
Qy 2696 GTTTCAGAGGATGGGTGAAACAGCTACAGATTGACTACAAAAATCACGAAATCATGACTG 2755
Db 2702 GTTTCAGAGGGTGGGTGAAGCAGTTGCAAATAGATTACAAAGGCAACGAAATAATGACGG 2761
Qy 2756 CGGCTGCATCGCAAGGACTTACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACG 2815
Db 2762 CAGCTGCCTCTCAAGGGCTGACCCGTAAAGGCGTGTATGCCGTTCGGTACAAGGTGAATG 2821
Qy 2816 AGAATCCACTCTACTCGCAGACTTCTGAGCACGTGAACGTGTTACTTACACGCACAGAAA 2875
Db 2822 AAAATCCCCTGTACGCACCCACCTCAGAACATGTGAACGTCCTACTGACCCGCACGGAGG 2881
Qy 2876 AACGCATTGTCTGGAAGACGCTAGCTGGTGATCCCTGGATAAAGACACTTACAGCTAAAT 2935
Db 2882 ACCGTATCGTGTGGAAAACACTAGCCGGTGATCCATGGATAAAAACACTGACGGCCAAGT 2941
Qy 2936 ATCCCGGGGATTTCACGGCTTCATTGGACGACTGGCAGCGCGAACACGACGCCATTATGG 2995
Db 2942 ATCCTGGGAACTTCACTGCCACGATAGAGGAATGGCAAGCAGAGCATGATGCCATCATGA 3001
Qy 2996 CACGCGTTCTTGATAAGCCGCAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCT 3055
Db 3002 GGCACATCTTGGAGAGACCGGACCCTACCGACGTTTTCCAAAATAAGGCGAACGTGTGTT 3061
Qy 3056 GGGCGAAGGCTTTAGAGCCAGTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGT 3115
Db 3062 GGGCCAAGGCTTTGGTGCCGGTACTGAAGACTGCAGGCATAGACATGACCACTGAACAAT 3121
Qy 3116 GGGAGACGTTGCACCCATTCAAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACT 3175
Db 3122 GGAACACTGTGGATTACTTCGAAACGGACAAAGCTCACTCAGCAGAGATAGTATTGAACC 3181
Qy 3176 TCTTTTGCACCAGGTTCTTTGGAGTAGACCTGGACAGTGGGTTATTTTCCGCTCCTACCG 3235
Db 3182 AACTATGCGTGAGGTTCTTTGGACTCGACCTGGACTCCGGTCTATTTTCTGCACCCACTG 3241

```
Qy       3236 TCGCACTTACTTACAGGGATCAGCACTGGGATAACTCGCCAGGGAAGAACATGTATGGGC 3295
Db       3242 TTCCGTTATCCATTAGGAATAATCACTGGGATAATTCCCCGTCGCCTAACATGTACGGGT 3301

Qy       3296 TTAATAGAGAGGTAGCAAAGGAGTTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTG 3355
Db       3302 TGAATAAAGAAGTGGTCCGCCAGCTCTCCCGCAGGTACCCACAACTGCCTCGAGCAGTTG 3361

Qy       3356 ACACAGGCAGGGTAGCTGATATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTA 3415
Db       3362 CCACCGGAAGAGTCTATGACATGAATACTGGCACGCTGCGCAATTATGATCCGCGCATAA 3421

Qy       3416 ATGTGGTTCCATTAAATCGCCGGTTGCCCCACTCGTTGATCGTTGACCACAAAGGACAGG 3475
Db       3422 ATCTAGTACCTGTGAACAGAAGACTGCCTCATGCTTTAGTCCTCCACCATAATGAACACC 3481

Qy       3476 GTACAACTGATCACAGCGGATTCCTATCTAAGATGAAGGGCAAATCTGTGTTGGTGATCG 3535
Db       3482 CACAGAGTGACTTTTCTTCATTCGTCAGCAAACTGAAGGGCAGAACTGTCTTGGTGGTCG 3541

Qy       3536 GCGATCCTATCAGCATTCCAGGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATA 3595
Db       3542 GGGAGAAGTTGTCCGTCCCAGGCAAAAAGGTCGACTGGTTGTCAGACCAGCCTGAGGCTA 3601

Qy       3596 CCATCAGGTGTGATCTCGATTTGGGAATACCTAGCCATGTCGGTAAATATGACATTATCT 3655
Db       3602 CCTTTAGAGCTCGGCTGGATTTAGGTATCCCAGGTGACGTGCCCAAATACGACATTGTAT 3661

Qy       3656 TTGTCAATGTTAGGACCCCGTACAGGAACCATCACTACCAACAGTGCGAGGATCACGCTA 3715
Db       3662 TTATTAACGTGAGGACTCCATATAAATACCATCATTATCAGCAGTGTGAAGACCACGCCA 3721

Qy       3716 TCCACCACAGCATGCTAACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTG 3775
Db       3722 TTAAGCTTAGTATGTTGACCAAGAAAGCTTGTCTGCATTTGAATCCCGGCGGAACCTGCG 3781

Qy       3776 TGGCTATAGGGTATGGGCTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCAC 3835
Db       3782 TCAGCATAGGTTATGGTTACGCTGACAGGGCCAGCGAGAGCATCATTGGTGCTATAGCGC 3841

Qy       3836 GCTCATTTAGGTTTACCCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTC 3895
Db       3842 GGCAGTTCAAGTTCTCCCGGGTATGCAAACCGAAATCCTCACATGAAGAGACAGAAGTAC 3901

Qy       3896 TCTTCGTGTTCTTCGGCAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTG 3955
Db       3902 TGTTTGTATTCATTGGGTACGATCGCAAGGCCCGTACGCACAATCCTTACAAGCTTTCAT 3961

Qy       3956 TAGTGCTTGACAACATCTATCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGT 4015
Db       3962 CTACCTTGACCAACATCTATACAGGTTCCAGACTCCACGAAGCCGGATGCGCACCCTCAT 4021

Qy       4016 ACAGAGTGATCAGAGGTGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTA 4075
Db       4022 ATCATGTGGTGCGAGGGGATATTGCCACGGCCACCGAAGGAGTGATCATAAATGCTGCCA 4081

Qy       4076 ATAGCAAAGGTCAACCAGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTG 4135
Db       4082 ACAGCAAAGGACAACCTGGCGGAGGGGTGTGCGGAGCGCTGTATAAGAAATTCCCGGAAA 4141

Qy       4136 CTTTTGATAGACAGCCAATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACC---GC 4192
Db       4142 GCTTCGATTTACAGCCGATCGAAGTGGGAAAAGCGCGACTGGTCAAAGGTGTAACTAAAC 4201

Qy       4193 TCATCATACATGCTGTAGGACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTA 4252
Db       4202 ATATTATTCATGCCGTAGGGCCCAACTTCAACAAAGTTTCGGAAGTTGAAGGGGACAAAC 4261

Qy       4253 AGCTCGCAGCTGCCTACATGAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAA 4312
Db       4262 AGTTGGCAGAGGCTTATGAGTCCATCGCTAAAATTGTCAACGATAACAATTACAAGTCAG 4321

Qy       4313 TATCAGTACCGCTACTGTCAACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAAT 4372

Db       4322 TAGCGATTCCACTGTTGTCCACCGGCATCTTTTCCGGGAACAAAGATCGACTAACCCAAT 4381

Qy       4373 CATTGCATCACCTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCT 4432
Db       4382 CATTGAACCATTTGCTGACAGCTTTAGACACCACTGATGCAGATGTAGCCATATACTGCA 4441

Qy       4433 TGGATAAACAATGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAA 4492
Db       4442 GGGACAAGAAATGGGAAATGACTCTCAAGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGG 4501

Qy       4493 TTCTGGA----------------TGATGACAAGCCAGTAGACATTGACTTGGTCAGGGTCC 4537
Db       4502 AGATATGCATATCAGACGACTCTTCGGTGACAGAACCGGATGCAGAGCTGGTGAGGGTAC 4561

Qy       4538 ACCCAAACAGCTCTTTGGCAGGCAGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTATT 4597
Db       4562 ATCCGAAGAGTTCTTTGGCTGGAAGGAAGGGCTACAGCACAAGTGATGGCAAGACTTTCT 4621

Qy       4598 CATACCTGGAAGGTACACGATTCCATCAGACCGCCAAGGACATTGCCGAAATCCATGCAA 4657
Db       4622 CATATTTGGAAGGGACCAAATTTCACCAGGCGGCCAAGGATATAGCAGAAATTAATGCCA 4681

Qy       4658 TGTGGCCCAACAAATCTGAGGCTAATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTA 4717
Db       4682 TGTGGCCAGTTGCAACGGAGGCCAATGAGCAAGTATGCATGTATATCCTCGGTGAAAGCA 4741

Qy       4718 TGTCCAGCATCCGCTCCAAATGCCCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACA 4777
Db       4742 TGAGCAGCATTAGGTCGAAATGCCCCGTCGAGGAGTCGGAAGCCTCCACACCACCTAGCA 4801

Qy       4778 CACTTCCATGCCTGTGTAATTACGCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTG 4837
Db       4802 CGCTGCCTTGCTTGTGCATCCATGCTATGACTCCAGAAAGAGTACAACGCCTAAAAGCCT 4861

Qy       4838 CGAAGAAAGAACAGTTCGCCGTATGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAG 4897
Db       4862 CACGTCCAGAACAAATTACTGTGTGCTCATCCTTTCCATTGCCGAAGTATAGAATCACTG 4921

Qy       4898 GCGTGCAGAAGCTACAGTGCAGCAAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTG 4957
Db       4922 GTGTGCAGAAGATCCAGTGCTCCCAGCCTATACTGTTCTCACCGAAGGTGCCTGCGTACA 4981

Qy       4958 TACACCCCAGGAAGTAC 4974
Db       4982 TTCATCCACGGAAGTAC 4998
```

```
RESULT 11
EEVNSPEPA
LOCUS       EEVNSPEPA              11447 bp ss-RNA     linear   VRL 25-MAY-1999
DEFINITION  Venezuelan equine encephalitis virus, complete genome.
ACCESSION   L01442
VERSION     L01442.2  GI:4887231
KEYWORDS    .
SOURCE      Venezuelan equine encephalitis virus
  ORGANISM  Venezuelan equine encephalitis virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
            Alphavirus; VEEV complex.
REFERENCE   1  (bases 1 to 11447)
  AUTHORS   Kinney,R.M., Johnson,B.J., Brown,V.L. and Trent,D.W.
  TITLE     Nucleotide sequence of the 26 S mRNA of the virulent Trinidad
            donkey strain of Venezuelan equine encephalitis virus and deduced
            sequence of the encoded structural proteins
  JOURNAL   Virology 152 (2), 400-413 (1986)
  MEDLINE   86263392
   PUBMED   3088830
REFERENCE   2  (bases 1 to 11447)
  AUTHORS   Johnson,B.J., Kinney,R.M., Kost,C.L. and Trent,D.W.
  TITLE     Molecular determinants of alphavirus neurovirulence: nucleotide and
            deduced protein sequence changes during attenuation of Venezuelan
            equine encephalitis virus
  JOURNAL   J. Gen. Virol. 67 (Pt 9), 1951-1960 (1986)
```

```
  MEDLINE   86306669
   PUBMED   3755750
REFERENCE   3  (bases 1 to 11447)
  AUTHORS   Kinney,R.M., Johnson,B.J., Welch,J.B., Tsuchiya,K.R. and Trent,D.W.
  TITLE     The full-length nucleotide sequences of the virulent Trinidad
            donkey strain of Venezuelan equine encephalitis virus and its
            attenuated vaccine derivative, strain TC-83
  JOURNAL   Virology 170 (1), 19-30 (1989)
  MEDLINE   89243175
   PUBMED   2524126
REFERENCE   4  (bases 1 to 11447)
  AUTHORS   Kinney,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-1993) Molecular Virology, Centers for Disease
            Control and Prevention, PO Box 2087, Fort Collins, CO 80522, USA
REFERENCE   5  (bases 1 to 11447)
  AUTHORS   Kinney,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-1999) Molecular Virology, Centers for Disease
            Control and Prevention, PO Box 2087, Fort Collins, CO 80522, USA
  REMARK    Sequence update by submitter
COMMENT     On May 25, 1999 this sequence version replaced gi:323711.
FEATURES             Location/Qualifiers
     source          1. .11447
                     /organism="Venezuelan equine encephalitis virus"
                     /mol_type="genomic RNA"
                     /strain="Trinidad donkey"
                     /isolate="donkey/Trinidad/1943"
                     /db_xref="taxon:11036"
     mRNA            1. .11447
                     /note="putative"
                     /function="genomic mRNA"
     gene            1. .7526
                     /gene="NS"
     5'UTR           1. .44
                     /gene="NS"
     CDS             45. .7526
                     /gene="NS"
                     /function="virus replication"
                     /note="possible incorporation of arginine, cysteine or
                     tryptophan at read through of UGA codon"
                     /codon_start=1
                     /transl_except=(pos:5682. .5684,aa:Arg)
                     /product="putative nonstructural polyprotein precursor"
                     /protein_id="AAC19321.2"
                     /db_xref="GI:4887232"
                     /translation="MEKVHVDIEEDSPFLRALQRSFPQFEVEAKQVTDNDHANARAFS
                     HLASKLIETEVDPSDTILDIGSAPARRMYSKHKYHCICPMRCAEDPDRLYKYATKLKK
                     NCKEITDKELDKKMKELAAVMSDPDLETETMCLHDDESCRYEGQVAVYQDVYAVDGPT
                     SLYHQANKGVRVAYWIGFDTTPFMFKNLAGAYPSYSTNWADETVLTARNIGLCSSDVM
                     ERSRRGMSILRKKYLKPSNNVLFSVGSTIYHEKRDLLRSWHLPSVFHLRGKQNYTCRC
                     ETIVSCDGYVVKRIAISPGLYGKPSGYAATMHREGFLCCKVTDTLNGERVSFPVCTYV
                     PATLCDQMTGILATDVSADDAQKLLVGLNQRIVVNGRTQRNTNTMKNYLLPVVAQAFA
                     RWAKEYKEDQEDERPLGLRDRQLVMGCCWAFRRHKITSIYKRPDTQTIIKVNSDFHSF
                     VLPRIGSNTLEIGLRTRIRKMLEEHKEPSPLITAEDVQEAKCAADEAKEVREAEELRA
                     ALPPLAADVEEPTLEADVDLMLQEAGAGSVETPRGLIKVTSYAGEDKIGSYAVLSPQA
                     VLKSEKLSCIHPLAEQVIVITHSGRKGRYAVEPYHGKVVVPEGHAIPVQDFQALSESA
                     TIVYNEREFVNRYLHHIATHGGALNTDEEYYKTVKPSEHDGEYLYDIDRKQCVKKELV
                     TGLGLTGELVDPPFHEFAYESLRTRPAAPYQVPTIGVYGVPGSGKSGIIKSAVTKKDL
                     VVSAKKENCAEIIRDVKKMKGLDVNARTVDSVLLNGCKHPVETLYIDEAFACHAGTLR
                     ALIAIIRPKKAVLCGDPKQCGFFNMMCLKVHFNHEICTQVFHKSISRRCTKSVTSVVS
                     TLFYDKKMRTTNPKETKIVIDTTGSTKPKQDDLILTCFRGWVKQLQIDYKGNEIMTAA
                     ASQGLTRKGVYAVRYKVNENPLYAPTSEHVNVLLTRTEDRIVWKTLAGDPWIKTLTAK
                     YPGNFTATIEEWQAEHDAIMRHILERPDPTDVFQNKANVCWAKALVPVLKTAGIDMTT
                     EQWNTVDYFETDKAHSAEIVLNQLCVRFFGLDLDSGLFSAPTVPLSIRNNHWDNSPSP
                     NMYGLNKEVVRQLSRRYPQLPRAVATGRVYDMNTGTLRNYDPRINLVPVNRRLPHALV
                     LHHNEHPQSDFSSFVSKLKGRTVLVVGEKLSVPGKMVDWLSDRPEATFRARLDLGIPG
                     DVPKYDIIFVNVRTPYKYHHYQQCEDHAIKLSMLTKKACLHLNPGGTCVSIGYGYADR
                     ASESIIGAIARQFKFSRVCKPKSSLEETEVLFVFIGYDRKARTHNPYKLSSTLTNIYT
                     GSRLHEAGCAPSYHVVRGDIATATEGVIINAANSKGQPGGGVCGALYKKFPESFDLQP
                     IEVGKARLVKGAAKHIIHAVGPNFNKVSEVEGDKQLAEAYESIAKIVNDNNYKSVAIP
                     LLSTGIFSGNKDRLTQSLNHLLTALDTTDADVAIYCRDKKWEMTLKEAVARREAVEEI
                     CISDDSSVTEPDAELVRVHPKSSLAGRKGYSTSDGKTFSYLEGTKFHQAAKDIAEINA
                     MWPVATEANEQVCMYILGESMSSIRSKCPVEESEASTPPSTLPCLCIHAMTPERVQRL
                     KASRPEQITVCSSFPLPKYRITGVQKIQCSQPILFSPKVPAYIHPRKYLVETPPVDET
                     PEPSAENQSTEGTPEQPPLITEDETRTRTPEPIIIEEEEEDSISLLSDGPTHQVLQVE
                     ADIHGPPSVSSSSWSIPHASDFDVDSLSILDTLEGASVTSGATSAETNSYFAKSMEFL
                     ARPVPAPRTVFRNPPHPAPRTRTPSLAPSRACSRTSLVSTPPGVNRVITREELEALTP
                     SRTPSRSVSRTSLVSNPPGVNRVITREEFEAFVAQQQRRFDAGAYIFSSDTGQGHLQQ
                     KSVRQTVLSEVVLERTELEISYAPRLDQEKEELLRKKLQLNPTPANRSRYQSRKVENM
                     KAITARRILQGLGHYLKAEGKVECYRTLHPVPLYSSSVNRAFSSPKVAVEACNAMLKE
                     NFPTVASYCIIPEYDAYLDMVDGASCCLDTASFCPAKLRSFPKKHSYLEPTIRSAVPS
                     AIQNTLQNVLAAATKRNCNVTQMRELPVLDSAAFNVECFKKYACNNEYWETFKENPIR
                     LTEENVVNYITKLKGPKAAALFAKTHNLNMLQDIPMDRFVMDLKRDVKVTPGTKHTEE
                     RPKVQVIQAADPLATAYLCGIHRELVRRLNAVLLPNIHTLFDMSAEDFDAIIAEHFQP
                     GDCVLETDIASFDKSEDDAMALTALMILEDLGVDAELLTLIEAAFGEISSIHLPTKTK
                     FKFGAMMKSGMFLTLFVNTVINIVIASRVLRERLTGSPCAAFIGDDNIVKGVKSDKLM
                     ADRCATWLNMEVKIIDAVVGEKAPYFCGGFILCDSVTGTACRVADPLKRLFKLGKPLA
                     ADDEHDDDRRRALHEESTRWNRVGILSELCKAVESRYETVGTSIIVMAMTTLASSVKS
                     FSYLRGAPITLYG"
     mat_peptide     45. .1649
                     /gene="NS"
                     /product="nonstructural protein nsP1"
                     /function="minus strand RNA synthesis; methyltransferase;
                     guanyltransferase"
     mat_peptide     1650. .4031
                     /gene="NS"
                     /product="nonstructural protein nsP2"
                     /function="RNA helicase; nonstructural proteinase;
                     necessary for subgenomic 26S mRNA synthesis"
     mat_peptide     4032. .5702
                     /gene="NS"
                     /product="nonstructural protein nsP3"
     mat_peptide     5703. .7523
                     /gene="NS"
                     /product="nonstructural protein nsP4"
                     /function="RNA polymerase"
     CDS             45. .5684
                     /gene="NS"
                     /codon_start=1
                     /product="putative nonstructural polyprotein precursor"
                     /protein_id="AAC19323.2"
                     /db_xref="GI:4887233"
                     /translation="MEKVHVDIEEDSPFLRALQRSFPQFEVEAKQVTDNDHANARAFS
                     HLASKLIETEVDPSDTILDIGSAPARRMYSKHKYHCICPMRCAEDPDRLYKYATKLKK
                     NCKEITDKELDKKMKELAAVMSDPDLETETMCLHDDESCRYEGQVAVYQDVYAVDGPT
                     SLYHQANKGVRVAYWIGFDTTPFMFKNLAGAYPSYSTNWADETVLTARNIGLCSSDVM
                     ERSRRGMSILRKKYLKPSNNVLFSVGSTIYHEKRDLLRSWHLPSVFHLRGKQNYTCRC
                     ETIVSCDGYVVKRIAISPGLYGKPSGYAATMHREGFLCCKVTDTLNGERVSFPVCTYV
                     PATLCDQMTGILATDVSADDAQKLLVGLNQRIVVNGRTQRNTNTMKNYLLPVVAQAFA
                     RWAKEYKEDQEDERPLGLRDRQLVMGCCWAFRRHKITSIYKRPDTQTIIKVNSDFHSF
                     VLPRIGSNTLEIGLRTRIRKMLEEHKEPSPLITAEDVQEAKCAADEAKEVREAEELRA
                     ALPPLAADVEEPTLEADVDLMLQEAGAGSVETPRGLIKVTSYAGEDKIGSYAVLSPQA
                     VLKSEKLSCIHPLAEQVIVITHSGRKGRYAVEPYHGKVVVPEGHAIPVQDFQALSESA
                     TIVYNEREFVNRYLHHIATHGGALNTDEEYYKTVKPSEHDGEYLYDIDRKQCVKKELV
                     TGLGLTGELVDPPFHEFAYESLRTRPAAPYQVPTIGVYGVPGSGKSGIIKSAVTKKDL
                     VVSAKKENCAEIIRDVKKMKGLDVNARTVDSVLLNGCKHPVETLYIDEAFACHAGTLR
                     ALIAIIRPKKAVLCGDPKQCGFFNMMCLKVHFNHEICTQVFHKSISRRCTKSVTSVVS
                     TLFYDKKMRTTNPKETKIVIDTTGSTKPKQDDLILTCFRGWVKQLQIDYKGNEIMTAA
                     ASQGLTRKGVYAVRYKVNENPLYAPTSEHVNVLLTRTEDRIVWKTLAGDPWIKTLTAK
                     YPGNFTATIEEWQAEHDAIMRHILERPDPTDVFQNKANVCWAKALVPVLKTAGIDMTT
                     EQWNTVDYFETDKAHSAEIVLNQLCVRFFGLDLDSGLFSAPTVPLSIRNNHWDNSPSP
                     NMYGLNKEVVRQLSRRYPQLPRAVATGRVYDMNTGTLRNYDPRINLVPVNRRLPHALV
                     LHHNEHPQSDFSSFVSKLKGRTVLVVGEKLSVPGKMVDWLSDRPEATFRARLDLGIPG
                     DVPKYDIIFVNVRTPYKYHHYQQCEDHAIKLSMLTKKACLHLNPGGTCVSIGYGYADR
                     ASESIIGAIARQFKFSRVCKPKSSLEETEVLFVFIGYDRKARTHNPYKLSSTLTNIYT
                     GSRLHEAGCAPSYHVVRGDIATATEGVIINAANSKGQPGGGVCGALYKKFPESFDLQP
                     IEVGKARLVKGAAKHIIHAVGPNFNKVSEVEGDKQLAEAYESIAKIVNDNNYKSVAIP
                     LLSTGIFSGNKDRLTQSLNHLLTALDTTDADVAIYCRDKKWEMTLKEAVARREAVEEI
                     CISDDSSVTEPDAELVRVHPKSSLAGRKGYSTSDGKTFSYLEGTKFHQAAKDIAEINA
                     MWPVATEANEQVCMYILGESMSSIRSKCPVEESEASTPPSTLPCLCIHAMTPERVQRL
                     KASRPEQITVCSSFPLPKYRITGVQKIQCSQPILFSPKVPAYIHPRKYLVETPPVDET
                     PEPSAENQSTEGTPEQPPLITEDETRTRTPEPIIIEEEEEDSISLLSDGPTHQVLQVE
                     ADIHGPPSVSSSSWSIPHASDFDVDSLSILDTLEGASVTSGATSAETNSYFAKSMEFL
                     ARPVPAPRTVFRNPPHPAPRTRTPSLAPSRACSRTSLVSTPPGVNRVITREELEALTP
                     SRTPSRSVSRTSLVSNPPGVNRVITREEFEAFVAQQQ"
```

```
     old_sequence    106. .107
                     /gene="NS"
                     /note="this correction changes amino acid nsP1-21-Thr to
                     nsP1-21-Ser, this 1992 data corrects the 1989 reference in
                     citation [3]"
                     /citation=[3]
                     /replace="gc"
     old_sequence    1533. .1534
                     /gene="NS"
                     /note="this correction results in changing amino acid
                     nsP1-497-Arg to nsP1-497-Ala, this 1992 data corrects the
                     sequence reported in the 1989 reference in citation [3]"
                     /citation=[3]
                     /replace="gc"
     old_sequence    1696
                     /gene="NS"
                     /note="this correction results in changing amino acid
                     nsP2-16-Asp to nsP2-16-Ala, this correction adds another
                     difference between the Trinidad donkey and vaccine TC-83
                     strains reported in citation [3]"
                     /citation=[3]
                     /replace="c"
     gene            7532. .11444
                     /gene="26S mRNA"
     mRNA            7532. .11444
                     /gene="26S mRNA"
                     /note="the structural proteins of the virus are translated
                     as a polyprotein precursor molecule from an intracellular,
                     26S mRNA species that is identical to the 3'-one-third
                     portion of the viral genomic mRNA; putative"
                     /function="intracellular, subgenomic viral mRNA species"

  Query Match             42.4%;  Score 2119.6;  DB 14;  Length 11447;
  Best Local Similarity   64.6%;  Pred. No. 0;
  Matches 3220;  Conservative    0;  Mismatches 1744;  Indels   24;  Gaps    3;

Qy         10 ACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTC 69
Db         30 AATTACCTACCCAAAATGGAGAAAGTTCACGTTGACATCGAGGAAGACAGCCCATTCCTC 89

Qy         70 AAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAAT 129
Db         90 AGAGCTTTGCAGCGGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCACTGATAAT 149

Qy        130 GACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTC 189
Db        150 GACCATGCTAATGCCAGAGCGTTTTCGCATCTGGCTTCAAAACTGATCGAAACGGAGGTG 209

Qy        190 GACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAAT 249
Db        210 GACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCCGCCCGCAGAATGTATTCTAAG 269

Qy        250 CACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGG 309
Db        270 CACAAGTATCATTGTATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGATTGTATAAG 329

Qy        310 TATGCAGAAAGACTTAAGAAAAGT------GACATTACCGACAAGAACATAGCCTCTAAG 363
Db        330 TATGCAACTAAGCTGAAGAAAAACTGTAAGGAAATAACTGATAAGGAATTGGACAAGAAA 389

Qy        364 GCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATG 423
Db        390 ATGAAGGAGCTCGCCGCCGTCATGAGCGACCCTGACCTGGAAACTGAGACTATGTGCCTC 449

Qy        424 CACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCA 483
Db        450 CACGACGACGAGTCGTGTCGCTACGAAGGGCAAGTCGCTGTTTACCAGGATGTATACGCG 509

Qy        484 GTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGG 543
Db        510 GTTGACGGACCGACAAGTCTCTATCACCAAGCCAATAAGGGAGTTAGAGTCGCCTACTGG 569

Qy        544 ATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTAC 603
Db        570 ATAGGCTTTGACACCACCCCTTTTATGTTTAAGAACTTGGCTGGAGCATATCCATCATAC 629

Qy        604 AACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCA 663
Db        630 TCTACCAACTGGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCCTATGCAGCTCT 689

Qy        664 GATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCT 723
Db        690 GACGTTATGGAGCGGTCACGTAGAGGGATGTCCATTCTTAGAAAGAAGTATTTGAAACCA 749

Qy        724 ACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTA 783
Db        750 TCCAACAATGTTCTATTCTCTGTTGGCTCGACCATCTACCACGAGAAGAGGGACTTACTG 809

Qy        784 CGTAGCTGGCATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGA 843
Db        810 AGGAGCTGGCACCTGCCGTCTGTATTTCACTTACGTGGCAAGCAAAATTACACATGTCGG 869

Qy        844 TGTGGGACCATTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGA 903
Db        870 TGTGAGACTATAGTTAGTTGCGACGGGTACGTCGTTAAAAGAATAGCTATCAGTCCAGGC 929

Qy        904 CTATACGGTAAAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGC 963
Db        930 CTGTATGGGAAGCCTTCAGGCTATGCTGCTACGATGCACCGCGAGGGATTCTTGTGCTGC 989

Qy        964 AAAGTCACAGATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCA 1023
Db        990 AAAGTGACAGACACATTGAACGGGGAGAGGGTCTCTTTTCCCGTGTGCACGTATGTGCCA 1049

Qy       1024 GCCACACTTTGCGATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCA 1083
Db       1050 GCTACATTGTGTGACCAAATGACTGGCATACTGGCAACAGATGTCAGTGCGGACGACGCG 1109

Qy       1084 CAAAAACTATTGGTTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAAT 1143
Db       1110 CAAAAACTGCTGGTTGGGCTCAACCAGCGTATAGTCGTCAACGGTCGCACCCAGAGAAAC 1169

Qy       1144 ACTAACACAATGCAGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCG 1203
Db       1170 ACCAATACCATGAAAAATTACCTTTGCCCGTAGTGGCCCAGGCATTTGCTAGGTGGGCA 1229

Qy       1204 CGTGAACATCGTGCCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTT 1263
Db       1230 AAGGAATATAAGGAAGATCAAGAAGATGAAAGGCCACTAGGACTACGAGATAGACAGTTA 1289

Qy       1264 ACTATGGGCTGCTGCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCT 1323
Db       1290 GTCATGGGGTGTTGTTGGGCTTTTAGAAGGCACAAGATAACATCTATTTATAAGCGCCCG 1349

Qy       1324 GGTACGCAAACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTT 1383
Db       1350 GATACCCAAACCATCATCAAAGTGAACAGCGATTTCCACTCATTCGTGCTGCCCAGGATA 1409

Qy       1384 ACCAGCCACGGGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACT 1443
Db       1410 GGCAGTAACACATTGGAGATCGGGCTGAGAACAAGAATCAGGAAAATGTTAGAGGAGCAC 1469

Qy       1444 GTCAAACCCGCACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAA 1503
Db       1470 AAGGAGCCGTCACCTCTCATTACCGCCGAGGACGTACAAGAAGCTAAGTGCGCAGCCGAT 1529

Qy       1504 GAAGCTGAAGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCT 1563
Db       1530 GAGGCTAAGGAGGTGCGTGAAGCCGAGGAGTTGCGCGCAGCTCTACCACCTTTGGCAGCT 1589

Qy       1564 GAAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGAGCA 1623
Db       1590 GATGTTGAGGAGCCCACTCTGGAAGCCGATGTCGACTTGATGTTACAAGAGGCTGGGGCC 1649

Qy       1624 GGTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAG 1683
```

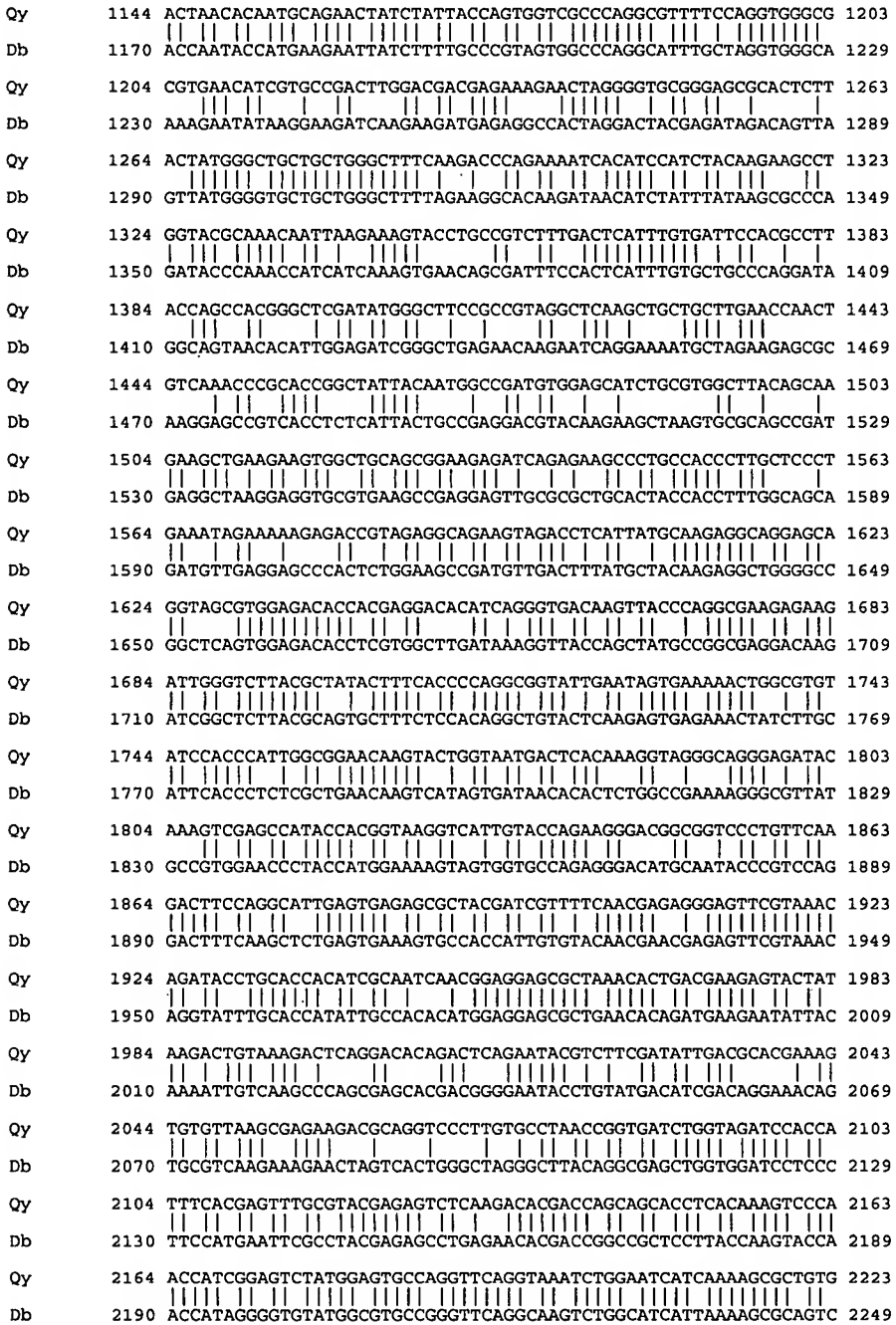

Db 1650 GGCTCAGTGGAGACACCTCGTGGCTTGATAAAGGTTACCAGCTACGCTGGCGAGGACAAG 1709
Qy 1684 ATTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGT 1743
Db 1710 ATCGGCTCTTACGCTGTGCTTTCTCCGCAGGCTGTACTCAAGAGTGAAAAATTATCTTGC 1769
Qy 1744 ATCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATAC 1803
Db 1770 ATCCACCCTCTCGCTGAACAAGTCATAGTGATAACACACTCTGGCCGAAAAGGGCGTTAT 1829
Qy 1804 AAAGTCGAGCCATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAA 1863
Db 1830 GCCGTGGAACCATACCATGGTAAAGTAGTGGTGCCAGAGGGACATGCAATACCCGTCCAG 1889
Qy 1864 GACTTCCAGGCATTGAGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAAC 1923
Db 1890 GACTTTCAAGCTCTGAGTGAAAGTGCCACCATTGTGTACAACGAACGTGAGTTCGTAAAC 1949
Qy 1924 AGATACCTGCACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTAT 1983
Db 1950 AGGTACCTGCACCATATTGCCACACATGGAGGAGCGCTGAACACTGATGAAGAATATTAC 2009
Qy 1984 AAGACTGTAAAGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAG 2043
Db 2010 AAAACTGTCAAGCCCAGCGAGCACGACGGCGAATACCTGTACGACATCGACAGGAAACAG 2069
Qy 2044 TGTGTTAAGCGAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCA 2103
Db 2070 TGCGTCAAGAAAGAACTAGTCACTGGGCTAGGGCTCACAGGCGAGCTGGTGGATCCTCCC 2129
Qy 2104 TTTCACGAGTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCA 2163
Db 2130 TTCCATGAATTCGCCTACGAGAGTCTGAGAACACGACCAGCCGCTCCTTACCAAGTACCA 2189
Qy 2164 ACCATCGGAGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTG 2223
Db 2190 ACCATAGGGGTGTATGGCGTGCCAGGATCAGGCAAGTCTGGCATCATTAAAAGCGCAGTC 2249
Qy 2224 ACTAAGAAAGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGAT 2283
Db 2250 ACCAAAAAAAGATCTAGTGGTGAGCGCCAAGAAAGAAAACTGTGCAGAAATTATAAGGGAC 2309
Qy 2284 GTAAGGAGGATGAGACGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAAT 2343
Db 2310 GTCAAGAAAATGAAAGGGCTGGACGTCAATGCCAGAACTGTGGACTCAGTGCTCTTGAAT 2369
Qy 2344 GGGGTTAAGCACCCCGTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGG 2403
Db 2370 GGATGCAAACACCCCGTAGAGACCCTGTATATTGACGAAGCTTTTGCTTGTCATGCAGGT 2429
Qy 2404 ACGCTGCTGGCACTGATTGCCATCGTCAAACCTAAGAAAGTGGTATTGTGCGGGGACCCA 2463
Db 2430 ACTCTCAGAGCGCTCATAGCCATTATAAGACCTAAAAAGGCAGTGCTCTGCGGGGATCCC 2489
Qy 2464 AAACAATGCGGCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGC 2523
Db 2490 AAACAGTGCGGTTTTTTTAACATGATGTGCCTGAAAGTGCATTTTAACCACGAGATTTGC 2549
Qy 2524 ACTGAAGTGTACCATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATCGTC 2583
Db 2550 ACACAAGTCTTCCACAAAAGCATCTCTCGCCGTTGCACTAAATCTGTGACTTCGGTCGTC 2609
Qy 2584 TCCACGCTCTTCTACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATCATC 2643
Db 2610 TCAACCTTGTTTTACGACAAAAAAATGAGAACGACGAATCCGAAAGAGACTAAGATTGTG 2669
Qy 2644 ATAGATACCACAGGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGA 2703
Db 2670 ATTGACACTACCGGCAGTACCAAACCTAAGCAGGACGATCTCATTCTCACTTGTTTCAGA 2729
Qy 2704 GGATGGGTGAAACAGCTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCTGCA 2763
Db 2730 GGGTGGGTGAAGCAGTTGCAAATAGATTACAAAGGCAACGAAATAATGACGGCAGCTGCC 2789
Qy 2764 TCGCAAGGACTTACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCA 2823
Db 2790 TCTCAAGGGCTGACCCGTAAAGGTGTGTATGCCGTTCGGTACAAGGTGAATGAAAATCCT 2849
Qy 2824 CTCTACTCGCAGACTTCTGAGCACGTGAACGTGTTACTTACACGCACAGAAAAACGCATT 2883
Db 2850 CTGTACGCACCCACCTCAGAACATGTGAACGTCCTACTGACCCGCACGGAGGACCGCATC 2909
Qy 2884 GTCTGGAAGACGCTAGCTGGTGATCCCTGGATAAAGACACTTACAGCTAAATATCCCGGG 2943
Db 2910 GTGTGGAAAACACTAGCCGGCGACCCATGGATAAAAACACTGACTGCCAAGTACCCTGGG 2969
Qy 2944 GATTTCACGGCTTCATTGGACGACTGGCAGCGCGAACACGACGCCATTATGGCACGCGTT 3003
Db 2970 AATTTCACTGCCACGATAGAGGAGTGGCAAGCAGAGCATGATGCCATCATGAGGCACATC 3029
Qy 3004 CTTGATAAGCCGCAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAG 3063
Db 3030 TTGGAGAGACCGGACCCTACCGACGTCTTCCAGAATAAGGCAAACGTGTGTTGGGCCAAG 3089
Qy 3064 GCTTTAGAGCCAGTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAGACG 3123
Db 3090 GCTTTAGTGCCGGTGCTGAAGACCGCTGGCATAGACATGACCACTGAACAATGGAACACT 3149
Qy 3124 TTGCACCCATTCAAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTTTGC 3183
Db 3150 GTGGATTATTTTGAAACGGACAAAGCTCACTCAGCAGAGATAGTATTGAACCAACTATGC 3209
Qy 3184 ACCAGGTTCTTTGGAGTAGACCTGGACAGTGGGTTATTTTCCGCTCCTACCGTCGCACTT 3243
Db 3210 GTGAGGTTCTTTGGACTCGATCTGGACTCCGGTCTATTTTCTGCACCCACTGTTCCGTTA 3269
Qy 3244 ACTTACAGGGATCAGCACTGGGATAACTCGCCAGGGAAGAACATGTATGGGCTTAATAGA 3303
Db 3270 TCCATTAGGAATAATCACTGGGATAACTCCCCGTCGCCTAACATGTACGGGCTGAATAAA 3329
Qy 3304 GAGGTAGCAAAGGAGTTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACAGGC 3363
Db 3330 GAAGTGGTCCGTCAGCTCTCTCGCAGGTACCCACAACTGCCTCGGGCAGTTGCCACTGGA 3389
Qy 3364 AGGGTAGCTGATATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTGGTT 3423
Db 3390 AGAGTCTATGACATGAACACTGGTACACTGCGCAATTATGATCCGCGCATAAACCTAGTA 3449
Qy 3424 CCATTAAATCGCCGGTTGCCCCACTCGTTGATCGTTGACCACAAAGGACAGGGTACAACT 3483
Db 3450 CCTGTAAACAGAAGACTGCCTCATGCTTTAGTCCTCCACCATAATGAACACCCACAGAGT 3509
Qy 3484 GATCACAGCGGATTCCTATCTAAGATGAAGGGCAAATCTGTGTTGGTGATCGGCGATCCT 3543
Db 3510 GACTTTTCTTCATTCGTCAGCAAATTGAAGGGCAGAACTGTCCTGGTGGTCGGGGAAAAG 3569
Qy 3544 ATCAGCATTCCAGGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATACCATCAGG 3603
Db 3570 TTGTCCGTCCCAGGCAAAATGGTTGACTGGTTGTCAGACCGGCCTGAGGCTACCTTCAGA 3629
Qy 3604 TGTGATCTCGATTTGGGAATACCTAGCCATGTCGGTAAATATGACATTATCTTTGTCAAT 3663
Db 3630 GCTCGGCTGGATTTAGGCATCCCAGGTGATGTGCCCAAATATGACATAATATTTGTTAAT 3689
Qy 3664 GTTAGGACCCCGTACAGGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCACCAC 3723
Db 3690 GTGAGGACCCCATATAAATACCATCACTATCAGCAGTGTGAAGACCATGCCATTAAGCTT 3749
Qy 3724 AGCATGCTAACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCTATA 3783
Db 3750 AGCATGTTGACCAAGAAAGCTTGTCTGCATCTGAATCCCGGCGGAACCTGTGTCAGCATA 3809
Qy 3784 GGGTATGGGCTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCACGCTCATTT 3843
Db 3810 GGTTATGGTTACGCTGACAGGGCCAGCGAAAGCATCATTGGTGCTATAGCGCGGCAGTTC 3869

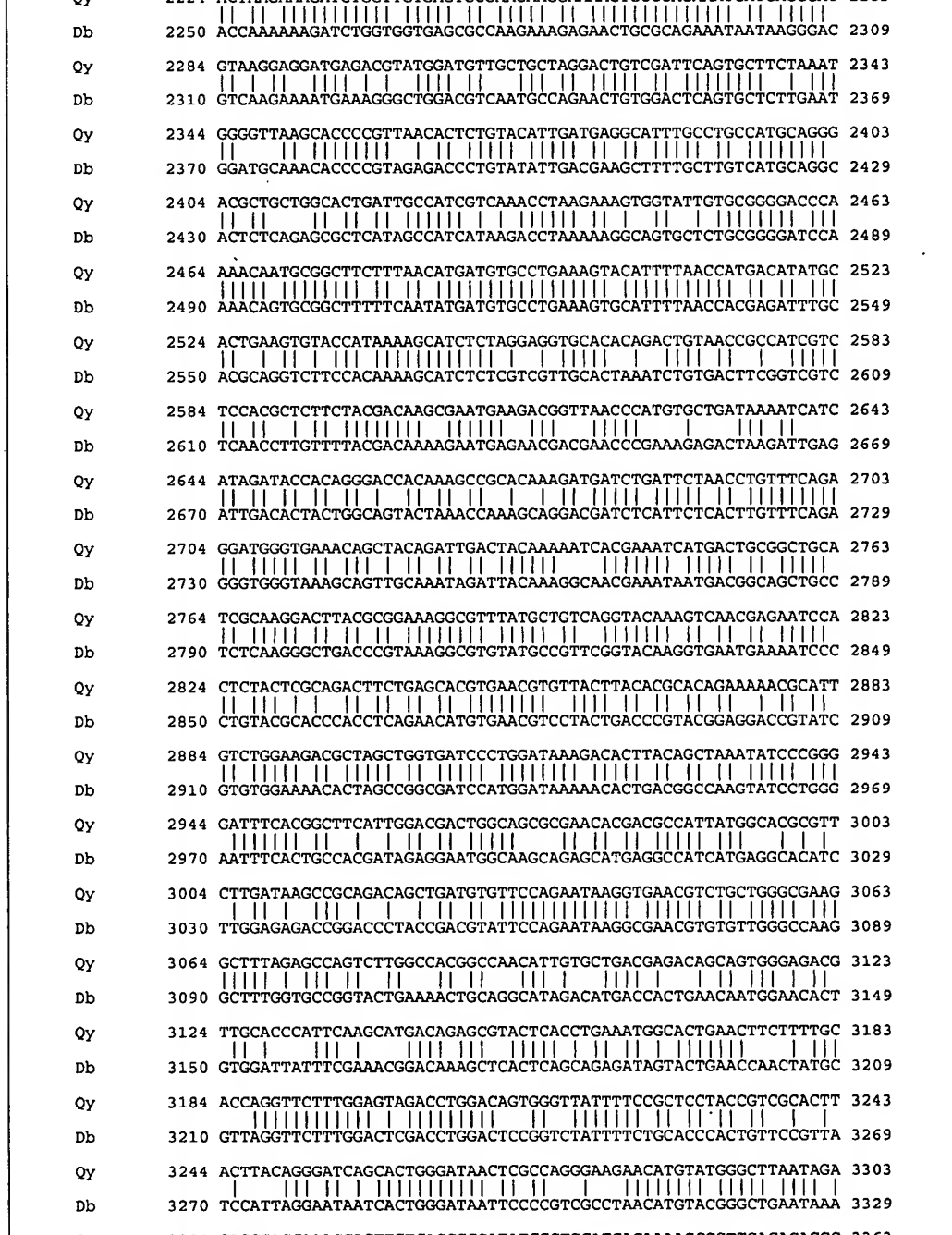
```
Qy       3844 AGGTTTACCCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTCGTG 3903
Db       3870 AAGTTTTCCCGGGTATGCAAACCGAAATCCTCACTTGAAGAGACGGAAGTTCTGTTTGTA 3929

Qy       3904 TTCTTCGGCAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTGCTT 3963
Db       3930 TTCATTGGGTACGATCGCAAGGCCCGTACGCACAATCCTTACAAGCTTTCATCAACCTTG 3989

Qy       3964 GACAACATCTATCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGTACAGAGTG 4023
Db       3990 ACCAACATTTATACAGGTTCCAGACTCCACGAAGCCGGATGTGCACCCTCATATCATGTG 4049

Qy       4024 ATCAGAGGTGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGCAAA 4083
Db       4050 GTGCGAGGGGATATTGCCACGGCCACCGAAGGAGTGATTATAAATGCTGCTAACAGCAAA 4109

Qy       4084 GGTCAACCAGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTTGAT 4143
Db       4110 GGACAACCTGGCGGAGGGGTGTGCGGAGCGCTGTATAAGAAATTCCCGGAAAGCTTCGAT 4169

Qy       4144 AGACAGCCAATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACC---GCTCATCATA 4200
Db       4170 TTACAGCCGATCGAAGTAGGAAAAGCGCGACTGGTCAAAGGTGCAGCTAAACATATCATT 4229

Qy       4201 CATGCTGTAGGACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTAAGCTCGCA 4260
Db       4230 CATGCCGTAGGACCAAACTTCAACAAAGTTTCGGAGGTTGAAGGTGACAAACAGTTGGCA 4289

Qy       4261 GCTGCCTACATGAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTA 4320
Db       4290 GAGGCTTATGAGTCCATCGCTAAGATTGTCAACGATAACAATTACAAGTCAGTAGCGATT 4349

Qy       4321 CCGCTACTGTCAACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCAT 4380
Db       4350 CCACTGTTGTCCACCGGCATCTTTTCCGGGAACAAAGATCGACTAACCCAATCATTGAAC 4409

Qy       4381 CACCTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAA 4440
Db       4410 CATTTGCTGACAGCTTTAGACACCACTGATGCAGATGTAGCCATATACTGCAGGGACAAG 4469

Qy       4441 CAATGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATTCTGGA- 4499
Db       4470 AAATGGGAAATGACTCTCAAGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGGAGATATGC 4529

Qy       4500 --------------TGATGACAAGCCAGTAGACATTGACTTGGTCAGGGTCCACCCAAAC 4545
Db       4530 ATATCCGACGACTCTTCAGTGACAGAACCTGATGCAGAGCTGGTGAGGGTGCATCCGAAG 4589

Qy       4546 AGCTCTTTGGCAGGCAGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTG 4605
Db       4590 AGTTCTTTGGCTGGAAGGAAGGGCTACAGCACAAGCGATGGCAAAACTTTCTCATATTTG 4649

Qy       4606 GAAGGTACACGATTCCATCAGACCGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCC 4665
Db       4650 GAAGGGACCAAGTTTCACCAGGCGGCCAAGGATATAGCAGAAATTAATGCCATGTGGCCC 4709

Qy       4666 AACAAATCTGAGGCTAATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGC 4725
Db       4710 GTTGCAACGGAGGCCAATGAGCAGGTATGCATGTATATCCTCGGAGAAAGCATGAGCAGT 4769

Qy       4726 ATCCGCTCCAAATGCCCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCA 4785
Db       4770 ATTAGGTCGAAATGCCCCGTCGAAGAGTCGGAAGCCTCCACACCACCTAGCACGCTGCCT 4829

Qy       4786 TGCCTGTGTAATTACGCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAA 4845
Db       4830 TGCTTGTGCATCCATGCCATGACTCCAGAAAGAGTACAGCGCCTAAAAGCCTCACGTCCA 4889

Qy       4846 GAACAGTTCGCCGTATGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAG 4905
Db       4890 GAACAAATTACTGTGTGCTCATCCTTTCCATTGCCGAAGTATAGAATCACTGGTGTGCAG 4949

Qy       4906 AAGCTACAGTGCAGCAAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCC 4965
Db       4950 AAGATCCAATGCTCCCAGCCTATATTGTTCTCACCGAAAGTGCCTGCGTATATTCATCCA 5009

Qy       4966 AGGAAGTA 4973
Db       5010 AGGAAGTA 5017
```

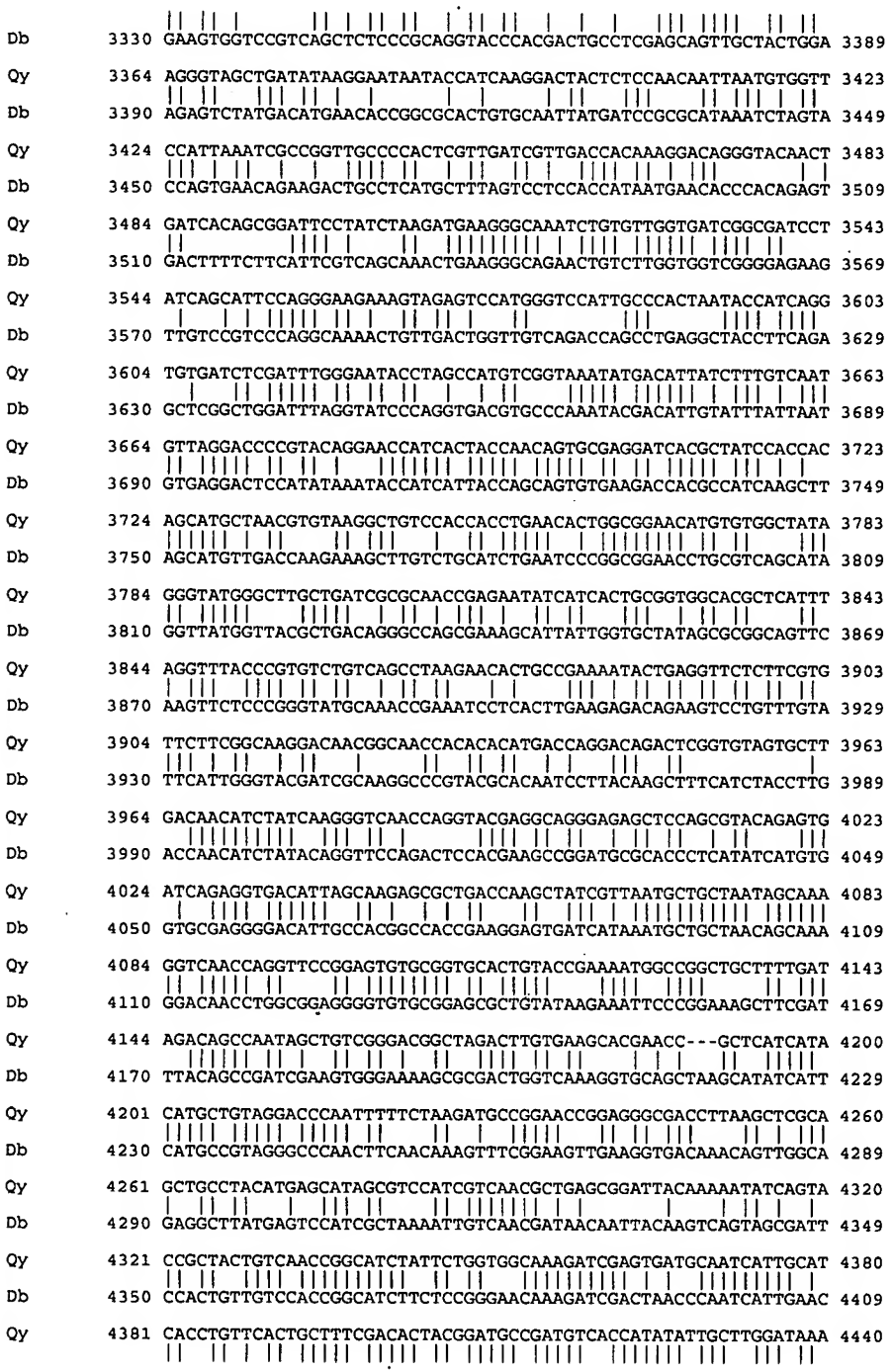
RESULT 12
AF004459

```
LOCUS       AF004459              11420 bp    RNA     linear   VRL 12-JUL-1999
DEFINITION  Venezuelan equine encephalitis virus strain 243937, complete
            genome.
ACCESSION   AF004459
VERSION     AF004459.2  GI:5442464
KEYWORDS    .
SOURCE      Venezuelan equine encephalitis virus
  ORGANISM  Venezuelan equine encephalitis virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
            Alphavirus; VEEV complex.
REFERENCE   1  (bases 8366 to 9182)
  AUTHORS   Powers,A.M., Oberste,M.S., Brault,A.C., Rico-Hesse,R.,
            Schmura,S.M., Smith,J.F., Kang,W., Sweeney,W.P. and Weaver,S.C.
  TITLE     Repeated emergence of epidemic/epizootic Venezuelan equine
            encephalitis from a single genotype of enzootic subtype ID virus
  JOURNAL   J. Virol. 71 (9), 6697-6705 (1997)
  MEDLINE   97404681
   PUBMED   9261393
REFERENCE   2  (bases 1 to 11420)
  AUTHORS   Wang,E., Barrera,R., Boshell,J., Ferro,C., Freier,J.E.,
            Navarro,J.C., Salas,R., Vasquez,C. and Weaver,S.C.
  TITLE     Genetic and phenotypic changes accompanying the emergence of
            epizootic subtype IC Venezuelan equine encephalitis viruses from an
            enzootic subtype ID progenitor
  JOURNAL   J. Virol. 73 (5), 4266-4271 (1999)
  MEDLINE   99214366
   PUBMED   10196323
REFERENCE   3  (bases 8366 to 9182)
  AUTHORS   Powers,A.M., Oberste,M.S., Brault,A.C., Rico-Hesse,R.,
            Schmura,S.M., Smith,J.F., Kang,W., Sweeney,W.P. and Weaver,S.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-1997) Pathology, Univ. Texas Med. Branch,
            Galveston, TX 77555-0609, USA
REFERENCE   4  (bases 1 to 11420)
  AUTHORS   Wang,E. and Weaver,S.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-1999) Pathology, Univ. Texas Med. Branch,
            Galveston, TX 77555-0609, USA
  REMARK    Sequence update by submitter
COMMENT     On Jul 12, 1999 this sequence version replaced gi:3849804.
FEATURES             Location/Qualifiers
     source          1. .11420
                     /organism="Venezuelan equine encephalitis virus"
                     /mol_type="genomic RNA"
                     /strain="243937"
                     /db_xref="taxon:11036"
     CDS             45. .7502
                     /note="contains nonstructural proteins 1, 2, 3 and 4;
                     in-frame termination codon is read through at low
                     frequency"
                     /codon_start=1
                     /transl_except=(pos:5658. .5660,aa:OTHER)
                     /product="nonstructural polyprotein"
                     /protein_id="AAD43357.1"
                     /db_xref="GI:5442466"
                     /translation="MEKVHVDIEEDSPFLRALQRSFPQFEVEAKQVTDNDHANARAFS
                     HLASKLIETEVDPSDTILDIGSAPARRMYSKHKYHCICPMRCAEDPDRLYKYATKLKK
                     NCKEITDKELDKKMKELAAVMSDPDLETETMCLHDDESCRYEGQVAVYQDVYAVDGPT
                     SLYHQADKGVRVAYWIGFDTTPFMFKNLAGAYPSYSTNWADETVLTARNIGLCSSDVM
                     ERSRRGMSILRKKYLKPSNNVLFSVGSTIYHEKRDLLRSWHLPSVFHLRGKQNYTCRC
                     ETIVSCDGYVVKRIAISPGLYGKPSGYAATMHREGFLCCKVTDTLNGERVSFPVCTYV
                     PATLCDQMTGILATDVSADDAQKLLVGLNQRIVVNGRTQRNTNTMKNYLLPVVAQAFA
                     RWAKEYKEDQEDERPLGLRDRQLVMGCCWAFRRHKITSIYKRPDTQTIIKVNSDFHSF
```

```
                     VLPRIGSNTLEIGLRTRIRKMLEERKEPSPLITAEDVQEAKCAADEAKEVREAEELRA
                     ALPPLAADVEEPTLEADVDFMLQEAGAGSVETPRGLIKVTSYAGEDKIGSYAVLSPQA
                     VLKSEKLSCIHPLAEQVIVITHSGRKGRYAVEPYHGKVVVPEGHAIPVQDFQALSESA
                     TIVYNEREFVNRYLHHIATHGGALNTDEEYYKIVKPSEHDGEYLYDIDRKQCVKKELV
                     TGLGLTGELVDPPFHEFAYESLRTRPAAPYQVPTIGVYGVPGSGKSGIIKSAVTKKDL
                     VVSAKKENCAEIIRDVKKMKGLDVNARTVDSVLLNGCKHPVETLYIDEAFACHAGTLR
                     ALIAIIRPKKAVLCGDPKQCGFFNMMCLKVHFNHEICTQVFHKSISRRCTKSVTSVVS
                     TLFYDKRMRTTNPKETKIEIDTTGSTKPKQDDLILTCFRGWVKQLQIDYKGNEIMTAA
                     ASQGLTRKGVYAVRYKVNENPLYAPTSEHVNVLLTRTEDRIVWKTLAGDPWIKTLTAK
                     YPGNFTATIEEWQAEHEAIMRHILERPDPTDVFQNKANVCWAKALVPVLKTAGIDMTT
                     EQWNTVDYFETDKAHSAEIVLNQLCVRFFGLDLDSGLFSAPTVPLSIRNNHWDNSPSP
                     NMYGLNKEVVRQLSRRYPRLPRAVATGRVYDMNTGALCNYDPRINLVPVNRRLPHALV
                     LHHNEHPQSDFSSFVSKLKGRTVLVVGEKLSVPGKTVDWLSDQPEATFRARLDLGIPG
                     DVPKYDIVFINVRTPYKYHHYQQCEDHAIKLSMLTKKACLHLNPGGTCVSIGYGYADR
                     ASESIIGAIARQFKFSRVCKPKSSLBETEVLFVFIGYDRKARTHNPYKLSSTLTNIYT
                     GSRLHEAGCAPSYHVVRGDIATATEGVIINAANSKGQPGGGVCGALYKKFPESFDLQP
                     IEVGKARLVKGAAKHIIHAVGPNFNKVSEVEGDKQLAEAYESIAKIVNDNNYKSVAIP
                     LLSTGIFSGNKDRLTQSLNHLLTALDTTDADVAIYCRDKKWEMTLKEAVARREAVEEI
                     CISDDSSVTEPDAELVRVHPKSSLAGRKGYSTSDGKTFSYLEGTKFYQAAKDIAEINA
                     MWPVATEANEQVCMYILGESMSSIRSKCPVEESEASTPPSTLPCLCIHAMTPERVQRL
                     KASRPEQITVCSSFPLPKYRITGVQKIQCSQPILFSPKVPAYIHPRKYLVETPPVEEN
                     QSTEGTPEPAPVNVDATRTRMPEPIIIEEEEEDSISLLSDGPTHQVLQVEADIHGPPS
                     VSSSSWSIPHASDFDVDSLSILDTLEGTSVTSEAASAETNSYFASSMEFLARPVPAPR
                     TVFRNPPLPAPRTRTPSLAPCRASSRTSLVSTPPGVNRVITREELEALTPSRAPSRSA
                     SRTSLVSNPPGVNRVITREEFEAFVAQQQXRFDAGAYIFSSDTGQGHLQQKSVRQTVL
                     SEVVLERTELEISYAPRLDQEKEELLRKKLQLNPTPANRSRYQSRRVENMKAITARRI
                     LQGLGHYLKAEGKVECYRTLHPVPLYSSSVNRAFSSPKVAVEACNAMLKENFPTVASY
                     CIIPEYDAYLDMVDGASCCLDTASFCPAKLRSFPKKHSYLEPTIRSAVPSAIQNTLQN
                     VLAAATKRNCNVTQMRELPVLDSAAFNVECFKKYACNNEYWETFKENPIRLTEENVIN
                     YITKLKGPKAAALFAKTHNLNMLQDIPMDRFVMDLKRDVKVTPGTKHTEERPKVQVIQ
                     AADPLATAYLCGIHRELVRRLNAVLLPNIHTLFDMSAEDFDAIIAEHFQPGDCVLETD
                     IASFDKSEDDAMALTALMILEDLGVDAELLTLIEAAFGEISSIHLPTKTKFKFGAMMK
                     SGMFLTLFVNTVINIVIASRVLRERLTGSPCAAFIGDDNIVKGVKSDKLMADRCATWL
                     NMEVKIIDAVVGEKAPYFCGGFILCDTVTGTACRVADPLKRLFKLGKPLAADDEHDDD
                     RRRALHEESTRWNRVGILPELCKAVESRYETVGTSIIVMAMTTLASSVKSFSYLRGAP
                     ITLYG"
     CDS             7538. .11305
                     /note="contains capsid, E3, E2, 6K and E1 proteins"
                     /codon_start=1
                     /product="structural polyprotein"
                     /protein_id="AAC71998.2"
                     /db_xref="GI:5442465"
                     /translation="MFPFQPMYPMQPMPYRNPFAAPRRPWFPRTDPFLAMQVQELTRS
                     MANLTFKQRRDAPPEGPPAKKPKREAPQKQKRGGQGKKKKNQGKKKAKTGPPNPKVQS
                     GNKKKTNKKPGKRQRMVMKLESDKTFPIMLEGKINGYACVVGGKLFRPMHVEGKIDND
                     VLAALKTKKASKYDLEYADVPQNMRADTFKYTHEKPQGYYSWHHGAVQYENGRFTVPK
                     GVGAKGDSGRPILDNQGRVVAIVLGGVNEGSRTALSVVMWNEKGVTVKYTPENCEQWS
                     LVTTMCLLANVTFPCAQPPICYDRKPAETLAMLSVNVDNPGYDELLEAAVKCPGRKRR
                     STEELFKEYKLTRPYMARCIRCAVGSCHSPIAIEAVKSDGHDGYVRLQTSSQYGLDSS
                     GNLKGRTMRYDMHGTIEEIPLHQVSLHTSRPCHIVDGHGYFLLARCPAGDSITMEFKK
                     NSVTHSCSVPYEVKFNPVGRELYTHPPEHGAEQACQVYAHDAQNRGAYVEMHLPGSEV
                     DSSLVSLSGSSVTVTPPARTSALVECECGGTKISETINRAKQFSQCTKKEQCRAYRLQ
                     NDKWVYNSDKLPKAAGATLKGKLHVPFLLADGKCTVPLAPEPMITFGFRSVSLKLHPK
                     NPTYLTTRQLADEPHYTHELISEPAVRNFTVTEKGWEFVWGNHPPKRFWAQETAPGNP
                     HGLPHEVITHYYHRYPMSTILGLSICAAIVTVSVAASTWLFCKSRVSCLTPYRLTPNA
                     RMPLCLAVLCCARTARAETTWESLDHLWNNNQQMFWIQLLIPLAALIVVTRLLKCVCC
                     VVPFLVVAGAAGAGAYEHATTMPSQAGISYNTIVNRAGYAPLPISITPTKIKLIPTVN
                     LEYVTCHYKTGMDSPAIKCCGSQECTPTYRPDEQCKVFTGVYPFMWGGAYCFCDTENT
                     QVSKAYVMKSDDCLADHAEAYKAHTASVQAFLNITVGEHSIVTTVYVNGETPVNFNGV
                     KLTAGPLSTAWTPFDRKIVQYAGEIYNYDFPEYGAGQPGAFGDIQSRTVSSSDLYANT
                     NLVLQRPKAGAIHVPYTQAPSGFEQWKKDKAPSLKFTAPFGCEIYTNPIRAENCAVGS
                     IPLAFDIPDALFTRVSETPTLSAAECTLNECVYSSDFGGIATVKYSASKSGKCAVHVP
                     SGTATLKEAAVELTEQGSVTIHFSTANIHPEFRLQICTSYVTCKGDCHPPKDHIVTHP
                     QYHAQTFTAAVSKTAWTWLTSLLGGSAVIIIIGLVLATIVAMYVLTNQKHN"
BASE COUNT     3221 a    2821 c    2894 g    2484 t
ORIGIN

  Query Match           42.4%;  Score 2119;  DB 14;  Length 11420;
  Best Local Similarity  64.5%;  Pred. No. 0;
  Matches 3220;  Conservative    0;  Mismatches 1745;  Indels   24;  Gaps    3;

Qy         10 ACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTC 69
              | | | | | |||| ||||| | | |||||||||||| | || |  |||||||| |   ||

Db         30 AATTACCTACCCAAAATGGAGAAAGTTCACGTTGACATCGAGGAAGACAGCCCATTCCTC 89

Qy         70 AAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAAT 129
              | | || || |||| || || || |||||| | ||||| | |||||||||||| |||

Db         90 AGAGCTTTGCAACGGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCACTGATAAT 149

Qy        130 GACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTC 189
              |||||||| || |||||||||||||||||| |||| |||| || || || |  || ||

Db        150 GACCATGCTAACGCCAGAGCGTTTTCGCATCTGGCTTCAAAACTGATCGAAACGGAGGTG 209

Qy        190 GACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAAT 249
              ||||    | |   ||| | ||||||||||||||||||| | |         |||| ||

Db        210 GACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCCGCCCGCAGAATGTATTCTAAG 269

Qy        250 CACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGG 309
              ||    |||||||| |||| || |||| | | || ||||| |||||||||  |  |   |

Db        270 CATAAGTATCATTGCATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGATTGTACAAG 329

Qy        310 TATGCAGAAAGACTTAAGAAAAGT------GACATTACCGACAAGAACATAGCCTCTAAG 363
              ||||||    | || |||||||  |      || || || ||||||  |  | |   ||

Db        330 TATGCAACTAAGCTGAAGAAAAATTGCAAGGAAATAACTGACAAGGAATTGGACAAGAAA 389

Qy        364 GCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATG 423
                 |   || ||    | ||||||       || |||   || |||    || ||||  |

Db        390 ATGAAGGAGCTTGCCGCCGTCATGAGCGACCCTGACCTGGAAACTGAGACTATGTGCCTC 449

Qy        424 CACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCA 483
              |||   ||||   | ||| | |||   ||     || || || ||||| ||||| ||||||

Db        450 CACGACGACGAGTCATGTCGCTACGAAGGGCAAGTCGCTGTTTACCAGGATGTATACGCA 509

Qy        484 GTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGG 543
              ||  | | |||||||     |||| |||||  ||     || ||||||||        ||||||

Db        510 GTTGACGGACCGACAAGTCTCTATCACCAAGCCGACAAGGGAGTTAGAGTCGCCTACTGG 569

Qy        544 ATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTAC 603
              |||||||||||||| ||||||||||||   || ||| |||| ||  | || ||  | |||

Db        570 ATAGGCTTTGACACCACCCCTTTTATGTTTAAGAACTTGGCTGGAGCATATCCATCATAC 629

Qy        604 AACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCA 663
                 || |||||||| ||||| |  || ||     || |||||||| |||||  | | |||

Db        630 TCTACCAACTGGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCCTATGCAGCTCC 689

Qy        664 GATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCT 723
              ||  |   |||| |    | | || |  || || ||||||||||||     |  ||||

Db        690 GACGTCATGGAGCGGTCACGTAGAGGGATGTCCATTCTTAGGAAGAAGTATTTGAAACCA 749

Qy        724 ACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTA 783
              | |||||  || ||||||| ||||| || || ||||||   ||  | ||    |  |

Db        750 TCCAATAATGTCCTATTCTCTGTTGGCTCGACCATCTACCACGAGAAGAGGGACTTACTG 809

Qy        784 CGTAGCTGGCATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGA 843
              | |||||||| || ||     || || |||||    || |||   || | |||| || |

Db        810 AGGAGCTGGCACCTGCCGTCTGTATTTCACTTACGTGGCAAGCAAAATTACACATGTCGG 869

Qy        844 TGTGGGACCATTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGA 903
              |||| ||| || || || || || |||||||| | |||| ||| | ||||| |||||

Db        870 TGTGAGACTATAGTTAGTTGCGACGGGTACGTCGTTAAAAGAATAGCTATCAGTCCAGGC 929

Qy        904 CTATACGGTAAAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGC 963
              || || || ||  |    ||  ||  |||| ||||| |||||||| |||||| | |||

Db        930 CTGTATGGGAAGCCTTCAGGCTATGCTGCCACGATGCACCGCGAGGGATTCTTGTGCTGC 989

Qy        964 AAAGTCACAGATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCA 1023
              ||||| ||||| ||  ||  ||| |||||||| |||||| | ||||| ||||||||  |||

Db        990 AAAGTGACAGACACATTGAACGGGGAGAGGGTCTCTTTTCCCGTGTGCACGTATGTGCCA 1049

Qy       1024 GCCACACTTTGCGATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCA 1083
              || ||| | || || |||| || || |||||||| || || |||| ||| |||||

Db       1050 GCTACATTGTGTGACCAAATGACTGGCATACTGGCAACAGATGTCAGTGCGGACGACGCG 1109

Qy       1084 CAAAAACTATTGGTTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAAT 1143
              |||||||  |||||||||||||||  | || |||||||| ||| | || ||||||||

Db       1110 CAAAAACTGCTGGTTGGGCTCAACCAGCGCATAGTCGTCAACGGTCGCACCCAAAGAAAC 1169
```

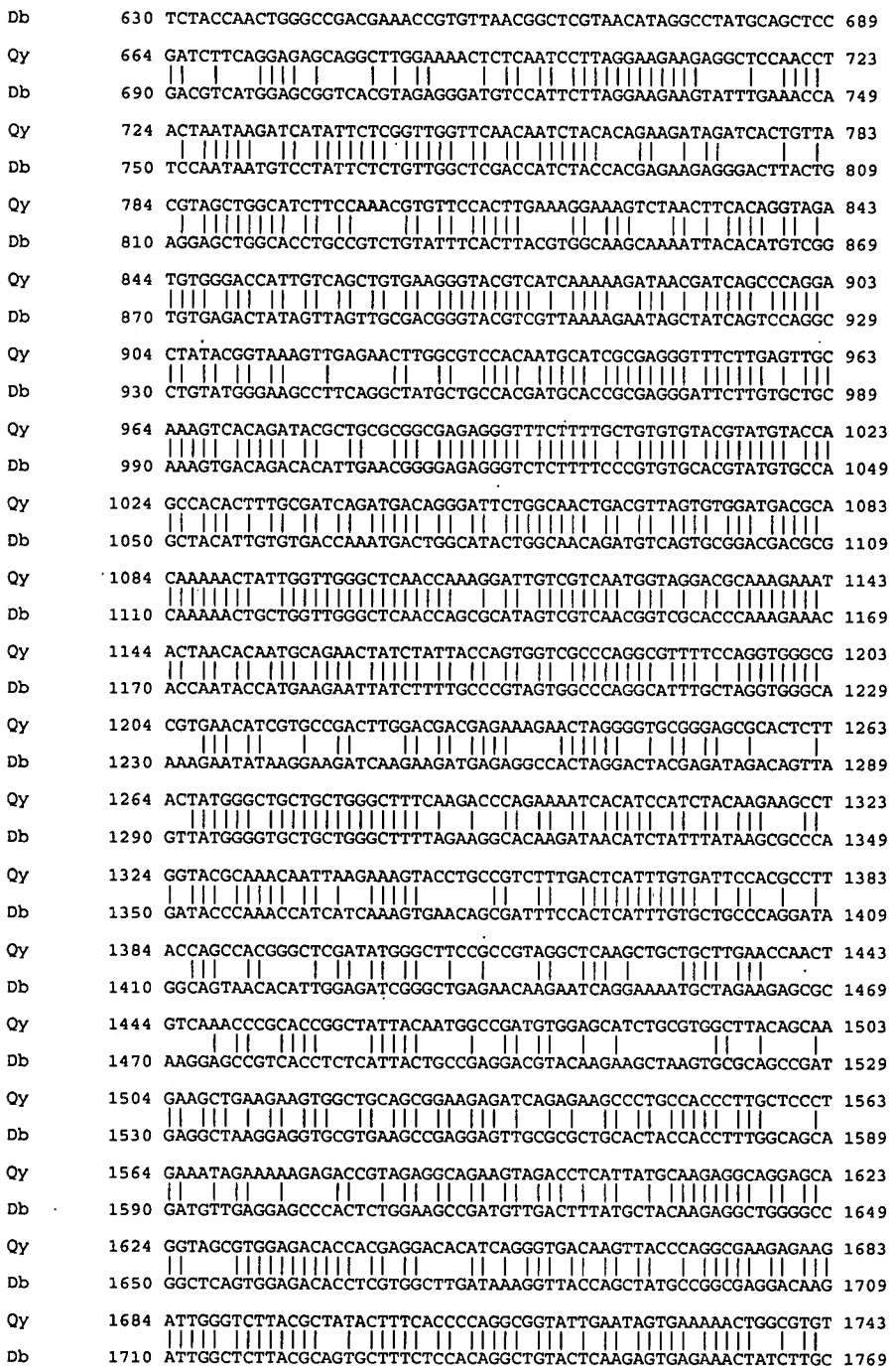


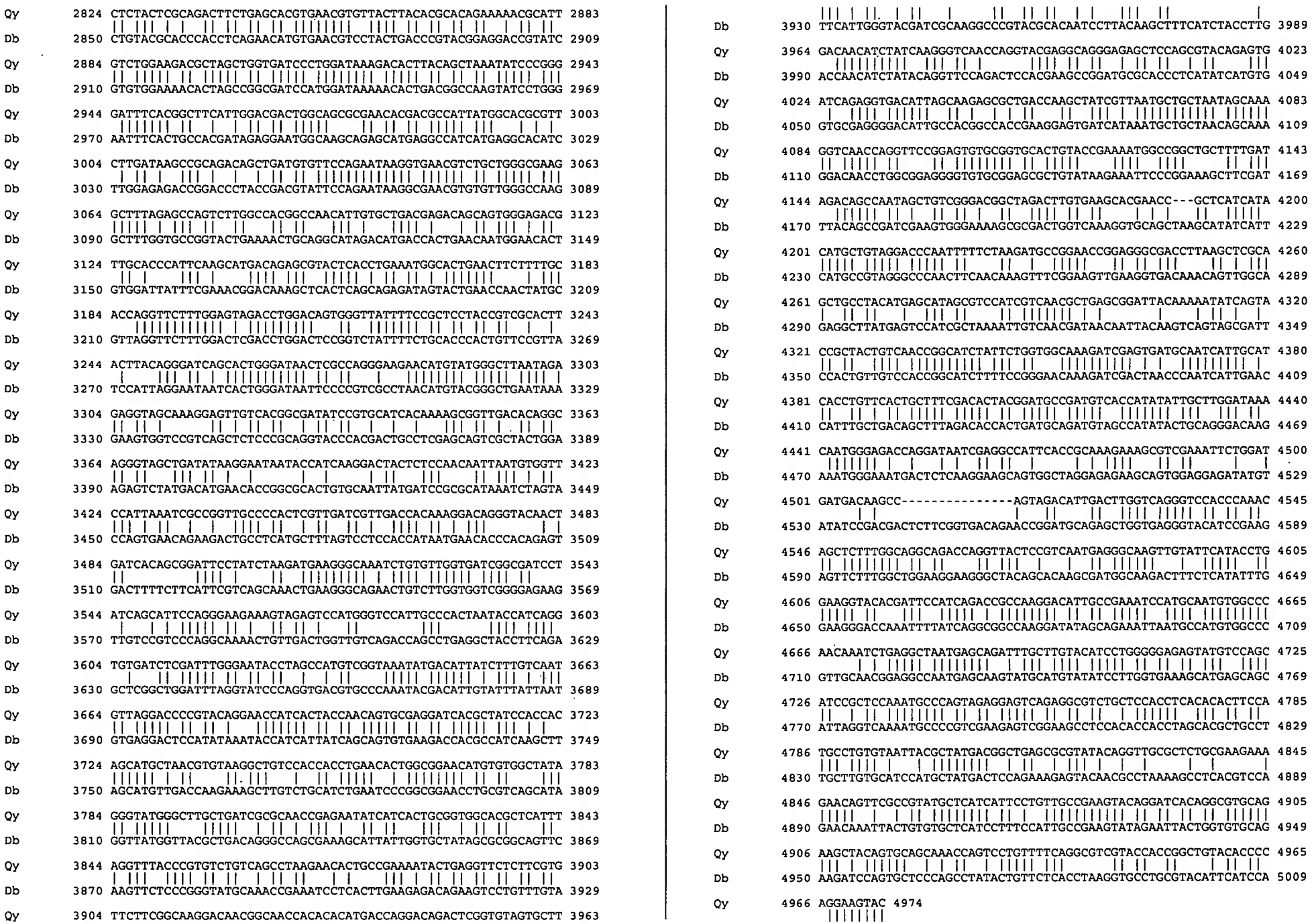




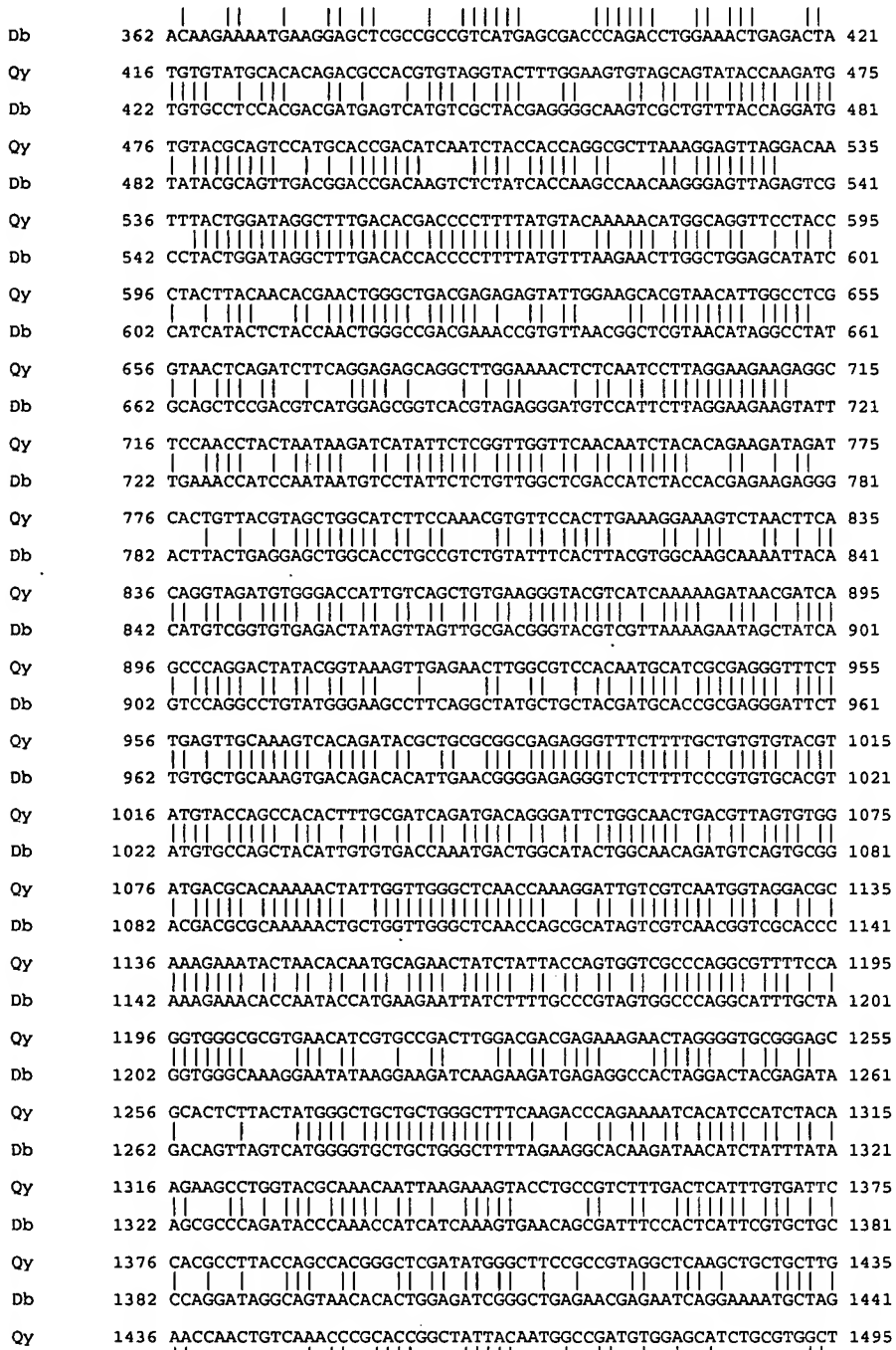
RESULT 13
VEU55360
LOCUS VEU55360 11420 bp RNA linear VRL 12-JUL-1999
DEFINITION Venezuelan equine encephalitis virus strain SH3, complete genome.
ACCESSION U55360
VERSION U55360.2 GI:5442468
KEYWORDS .
SOURCE Venezuelan equine encephalitis virus
ORGANISM Venezuelan equine encephalitis virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae; Alphavirus; VEEV complex.
REFERENCE 1 (bases 8366 to 10042)
AUTHORS Weaver,S.C., Salas,R., Rico-Hesse,R., Ludwig,G.V., Oberste,M.S., Boshell,J. and Tesh,R.B.
TITLE Re-emergence of epidemic Venezuelan equine encephalomyelitis in South America. VEE Study Group
JOURNAL Lancet 348 (9025), 436-440 (1996)
MEDLINE 96332169
PUBMED 8709783
REFERENCE 2 (bases 1 to 11420)
AUTHORS Wang,E., Barrera,R., Boshell,J., Ferro,C., Freier,J.E., Navarro,J.C., Salas,R., Vasquez,C. and Weaver,S.C.
TITLE Genetic and phenotypic changes accompanying the emergence of epizootic subtype IC Venezuelan equine encephalitis viruses from an enzootic subtype ID progenitor
JOURNAL J. Virol. 73 (5), 4266-4271 (1999)
MEDLINE 99214366
PUBMED 10196323
REFERENCE 3 (bases 8366 to 10042)
AUTHORS Weaver,S.C.

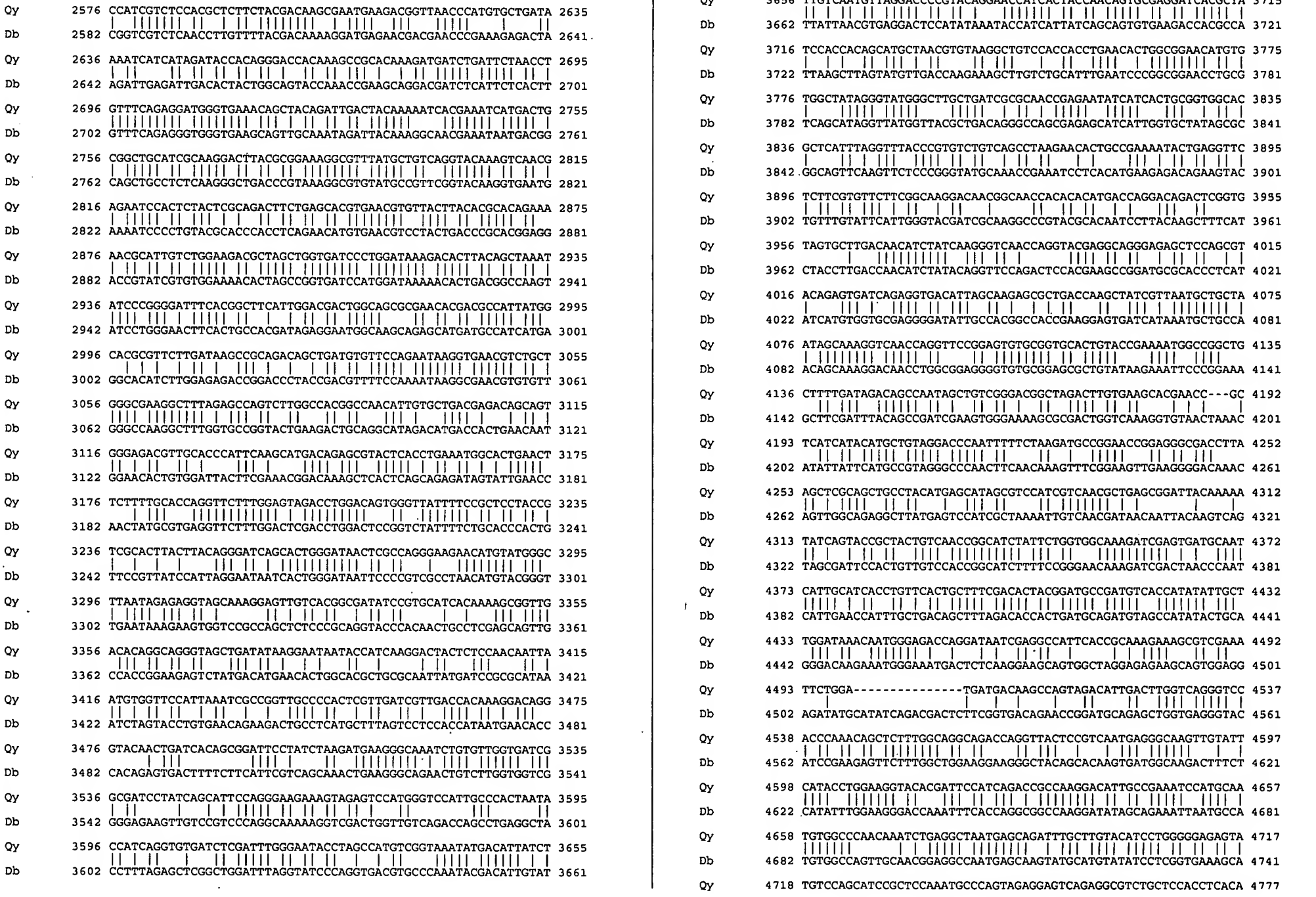
```
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-1996) Pathology, Univ. TX Med. Branch, 301
            University Ave., Galveston, TX 77555-0609, USA
REFERENCE   4  (bases 1 to 11420)
  AUTHORS   Wang,E. and Weaver,S.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-1999) Pathology, Univ. TX Med. Branch, 301
            University Ave., Galveston, TX 77555-0609, USA
  REMARK    Sequence update by submitter
COMMENT     On Jul 12, 1999 this sequence version replaced gi:3849806.
FEATURES             Location/Qualifiers
     source          1. .11420
                     /organism="Venezuelan equine encephalitis virus"
                     /mol_type="genomic RNA"
                     /strain="SH3"
                     /db_xref="taxon:11036"
     CDS             45. .7502
                     /note="contains nonstructural proteins 1, 2, 3 and 4;
                     in-frame termination codon is read through at low
                     frequency"
                     /codon_start=1
                     /transl_except=(pos:5658. .5660,aa:OTHER)
                     /product="nonstructural polyprotein"
                     /protein_id="AAD43358.1"
                     /db_xref="GI:5442470"
                     /translation="MEKVHVDIEEDSPFLRALQRSFPQFEVEAKQVTDNDHANARAFS
                     HLASKLIETEVDPSDTILDIGSAPARRMYSKHKYHCICPMRCAEDPDRLYKYATKLKK
                     NCKEITDKELDKKMKELAAVMSDPDLETETMCLHDDESCRYEGQVAVYQDVYAVDGPT
                     SLYHQADKGVRVAYWIGFDTTPFMFKNLAGAYPSYSTNWADETVLTARNIGLCSSDVM
                     ERSRRGMSILRKKYLKPSNNVLFSVGSTIYHEKRDLLRSWHLPSVFHLRGKQNYTCRC
                     ETIVSCDGYVVKRIAISPGLYGKPSGYAATMHREGFLCCKVTDTLNGERVSFPVCTYV
                     PATLCDQMTGILATDVSADDAQKLLVGLNQRIVVNGRTQRNTNTMKNYLLPVVAQAFA
                     RWAKEYKEDQEDERPLGLRDRQLVMGCCWAFRRHKITSIYKRPDTQTIIKVNSDFHSF
                     VLPRIGSNTLEIGLRTRIRKMLEERKEPSPLITAEDVQEAKCAADEAKEVREAEELRA
                     ALPPLAADVEEPTLEADVDFMLQEAGAGSVETPRGLIKVTSYAGEDKIGSYAVLSPQA
                     VLKSEKLSCIHPLAEQVIVITHSGRKGRYAVEPYHGKVVVPEGHAIPVQDFQALSESA
                     TIVYNEREFVNRYLHHIATHGGALNTDEEYYKIVKPSEHDGEYLYDIDRKQCVKKELV
                     TGLGLTGELVDPPFHEFAYESLRTRPAAPYQVPTIGVYGVPGSGKSGIIKSAVTKKDL
                     VVSAKKENCAEIIRDVKKMKGLDVNARTVDSVLLNGCKHPVETLYIDEAFACHAGTLR
                     ALIAIIRPKKAVLCGDPKQCGFFNMMCLKVHFNHEICTQVFHKSISRRCTKSVTSVVS
                     TLFYDKRMRTTNPKETKIEIDTTGSTKPKQDDLILTCFRGWVKQLQIDYKGNEIMTAA
                     ASQGLTRKGVYAVRYKVNENPLYAPTSEHVNVLLTRTEDRIVWKTLAGDPWIKTLTAK
                     YPGNFTATIEEWQAEHEAIMRHILERPDPTDVFQNKANVCWAKALVPVLKTAGIDMTT
                     EQWNTVDYFETDKAHSAEIVLNQLCVRFFGLDLDSGLFSAPTVPLSIRNNHWDNSPSP
                     NMYGLNKEVVRQLSRRYPRLPRAVATGRVYDMNTGALCNYDPRINLVPVNRRLPHALV
                     LHHNEHPQSDFSSFVSKLKGRTVLVVGEKLSVPGKTVDWLSDQPEATFRARLDLGIPG
                     DVPKYDIVFINVRTPYKYHHYQQCEDHAIKLSMLTKKACLHLNPGGTCVSIGYGYADR
                     ASESIIGAIARQFKFSRVCKPKSSLEETEVLFVFIGYDRKARTHNPYKLSSTLTNIYT
                     GSRLHEAGCAPSYHVVRGDIATATEGVIINAANSKGQPGGGVCGALYKKFPESFDLQP
                     IEVGKARLVKGAAKHIIHAVGPNFNKVSEVEGDKQLAEAYESIAKIVNDNNYKSVAIP
                     LLSTGIFSGNKDRLTQSLNHLLTALDTTDADVAIYCRDKKWEMTLKEAVARREAVEEI
                     CISDDSSVTEPDAELVRVHPKSSLAGRKGYSTSDGKTFSYLEGTKFYQAAKDIAEINA
                     MWPVATEANEQVCMYILGESMSSIRSKCPVEESEASTPPSTLPCLCIHAMTPERVQRL
                     KASRPEQITVCSSFPLPKYRITGVQKIQCSQPILFSPKVPAYIHPRKYLVETPPVEEN
                     QSTEGTPEPAPVNVDATRTRMPEPIIIEEEEEDSISLLSDGPTHQVLQVEADIHGPPS
                     VSSSSWSIPHASDFDVDSLSILDTLEGTSVTSEAASAETNSYFASSMEFLARPVPAPR
                     TVFRNPPHPAPRTRTPSLAPCRASSRTSLVSTPPGVNRVITREELEALTPSRAPSRSA
                     SRTSLVSNPPGVNRVITREEFEAFVAQQQXRFDAGAYTFSSDTGQGHLQQKSVRQTVL
                     SEVVLERTELEISYAPRLDQEKEELLRKKLQLNPTPANRSRYQSRRVENMKAITARRI
                     LQGLGHYLKAEGKVECYRTLHPVPLYSSSVNRAFSSPKVAVEACNAMLKENFPTVASY
                     CIIPEYDAYLDMVDGASCCLDTASFCPAKLRSFPKKHSYLEPTIRSAVPSAIQNTLQN
                     VLAAATKRNCNVTQMRELPVLDSAAFNVECFKKYACNNEYWETFKENPIRLTEENVIN
                     YITKLKGPKAAALFAKTHNLNMLQDIPMDRFVMDLKRDVKVTPGTKHTEERPKVQVIQ
                     AADPLATAYLCGIHRELVRRLNAVLLPNIHTLFDMSAEDFDAIIAEHFQPGDCVLETD
                     IASFDKSEDDAMALTALMILEDLGVDAELLTLIEAAFGEISSIHLPTKTKFKFGAMMK
                     SGMFLTLFVNTVINIVIASRVLRERLTGSPCAAFIGDDNIVKGVKSDKLMADRCATWL
                     NMEVKIIDAVVGEKAPYFCGGFILCDTVTGTACRVADPLKRLFKLGKPLAADDEHDDD
                     RRRALHEESTRWNRVGILPELCKAVESRYETVGTSIIVMAMTTLASSVKSFSYLRGAP
                     ITLYG"
     CDS             7538. .11305
                     /note="contains capsid, E3, E2, 6K and E1 proteins"
                     /codon_start=1
                     /product="structural polyprotein"
                     /protein_id="AAC71999.2"
                     /db_xref="GI:5442469"
                     /translation="MFPFQPMYPMQPMPYRNPFAAPRRPWFPRTDPFLAMQVQELTRS
                     MANLTFKQRRDAPPEGPPAKKPKREAPQKQKRGGQGKKKKNQGKKKAKTGPPNPKVQS
                     GNKKKTNKKPGKRQRMVMKLESDKTFPIMLEGKINGYACVVGGKLFRPMHVEGKIDND
                     VLAALKTKKASKYDLEYADVPQNMRADTFKYTHEKPQGYYSWHHGAVQYENGRFTVPK
                     GVGAKGDSGRPILDNQGRVVAIVLGGVNEGSRTALSVVMWNEKGVTVKYTPENCEQWS
                     LVTTMCLLANVTFPCAQPPICYDRKPAETLAMLSVNVDNPGYDELLEAAVKCPGRKRR
                     STEELFKEYKLTRPYMARCIRCAVGSCHSPIAIEAVKSDGHDGYVRLQTSSQYGLDSS
                     GNLKGRTMRYDMHGTIEEIPLHQVSLHTSRPCHIVDGHGYFLLARCPAGDSITMEFKK
                     NSVTHSCSVPYEVKFNPVGRELYTHPPEHGAEQACQVYAHDAQNRGAYVEMHLPGSEV
                     DSSLVSLSGSSVTVTPPARTSALVECECGGTKISETINRAKQFSQCTKKEQCRAYRLQ
                     NDKWVYNSDKLPKAAGATLKGKLHVPFLLADGKCTVPLAPEPMITFGFRSVSLKLHPK
                     NPTYLTTRQLADEPHYTHELISEPAVRNFTVTEKGWEFVWGNHPPKRFWAQETAPGNP
                     HGLPHEVITHYYHRYPMSTILGLSICAAIVTVSVAASTWLFCKSRVSCLTPYRLTPNA
                     RMPLCLAVLCCARTARAETTWESLDHLWNNNQQMFWIQLLIPLAALIVVTRLLKCVCC
                     VVPFLVVAGAAGAGAYEHATTMPSQAGISYNTIVNRAGYAPLPISITPTKIKLIPTVN
                     LEYVTCHYKTGMDSPAIKCCGSQECTPTYRPDEQCKVFTGVYPFMWGGAYCFCDTENT
                     QVSKAYVMKSDDCLADHAEAYKAHTASVQAFLNITVGEHSIVTTVYVNGETPVNFNGV
                     KLTAGPLSTAWTPFDRKIVQYAGEIYNYDFPEYGAGQPGAFGDIQSRTVSSSDLYANT
                     NLVLQRPKAGAIHVPYTQAPSGFEQWKKDKAPSLKFTAPFGCEIYTNPIRAENCAVGS
                     IPLAFDIPDALFTRVSETPTLSAAECTLNECVYSSDFGGIATVKYSASKSGKCAVHVP
                     SGTATLKEAAVELTEQGSVTIHFSTANIHPEFRLQICTSYVTCKGDCHPPRDHIVTHP
                     QYHAQTFTAAVSKTAWTWLTSLLGGSAVIIIIGLVLATIVAMYVLTNQKHN"
BASE COUNT     3220 a    2821 c    2895 g    2484 t
ORIGIN

  Query Match             42.4%;  Score 2119;  DB 14;  Length 11420;
  Best Local Similarity   64.5%;  Pred. No. 0;
  Matches 3220;  Conservative    0;  Mismatches 1745;  Indels   24;  Gaps    3;

Qy         10 ACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTC 69
              | | | | | |||| ||||| | | |||||||||||| | || | ||||||||  |   ||
Db         30 AATTACCTACCCAAAATGGAGAAAGTTCACGTTGACATCGAGGAAGACAGCCCATTCCTC 89

Qy         70 AAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAAT 129
              |   | || || ||||  || || || |||||| | ||||| | |||||||||||| |||
Db         90 AGAGCTTTGCAACGGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCACTGATAAT 149

Qy        130 GACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTC 189
              |||||||| || |||||||||||||||||| ||||  |||| || || || | || ||
Db        150 GACCATGCTAACGCCAGAGCGTTTTCGCATCTGGCTTCAAAACTGATCGAAACGGAGGTG 209

Qy        190 GACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAAT 249
              ||||    | |    ||| | ||||||||||||||||||| | |        |||| ||
Db        210 GACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCCGCCCGCAGAATGTATTCTAAG 269

Qy        250 CACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGG 309
              ||    |||||||| ||||| || |||| | | || ||||||||||||||| | |   |
Db        270 CATAAGTATCATTGCATCTGTCCGATGAGATGTGCGGAAGACCCGGACAGATTGTACAAG 329

Qy        310 TATGCAGAAAGACTTAAGAAAAGT------GACATTACCGACAAGAACATAGCCTCTAAG 363
              ||||||    |  || |||||||  |      || || || |||||| |  | | |  ||
Db        330 TATGCAACTAAGCTGAAGAAAAATTGCAAGGAAATAACTGACAAGGAATTGGACAAGAAA 389

Qy        364 GCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATG 423
                |  || ||    | ||||||      || |||   || |||    || |||| |
Db        390 ATGAAGGAGCTTGCCGCCGTCATGAGCGACCCTGACCTGGAAACTGAGACTATGTGCCTC 449

Qy        424 CACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCA 483
              |||  ||||   | ||| | |||   ||   || || || ||||| ||||| ||||||
Db        450 CACGACGACGAGTCATGTCGCTACGAAGGGCAAGTCGCTGTTTACCAGGATGTATACGCA 509

Qy        484 GTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGG 543
              ||  | | ||||||     |||| ||||| ||   || |||||||        ||||||
Db        510 GTTGACGGACCGACAAGTCTCTATCACCAAGCCGACAAGGGAGTTAGAGTCGCCTACTGG 569

Qy        544 ATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTAC 603
              |||||||||||||| ||||||||||||   || ||| |||| ||  | || ||  | |||
Db        570 ATAGGCTTTGACACCACCCCTTTTATGTTTAAGAACTTGGCTGGAGCATATCCATCATAC 629

Qy        604 AACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCA 663
                 || |||||||| |||||  |  || ||    || |||||||| |||||  | | |||
```


Db 630 TCTACCAACTGGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCCTATGCAGCTCC 689
Qy 664 GATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCT 723
Db 690 GACGTCATGGAGCGGTCACGTAGAGGGATGTCCATTCTTAGGAAGAAGTATTTGAAACCA 749
Qy 724 ACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTA 783
Db 750 TCCAATAATGTCCTATTCTCTGTTGGCTCGACCATCTACCACGAGAAGAGGGACTTACTG 809
Qy 784 CGTAGCTGGCATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGA 843
Db 810 AGGAGCTGGCACCTGCCGTCTGTATTTCACTTACGTGGCAAGCAAAATTACACATGTCGG 869
Qy 844 TGTGGGACCATTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGA 903
Db 870 TGTGAGACTATAGTTAGTTGCGACGGGTACGTCGTTAAAAGAATAGCTATCAGTCCAGGC 929
Qy 904 CTATACGGTAAAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGC 963
Db 930 CTGTATGGGAAGCCTTCAGGCTATGCTGCCACGATGCACCGCGAGGGATTCTTGTGCTGC 989
Qy 964 AAAGTCACAGATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCA 1023
Db 990 AAAGTGACAGACACATTGAACGGGGAGAGGGTCTCTTTTCCCGTGTGCACGTATGTGCCA 1049
Qy 1024 GCCACACTTTGCGATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCA 1083
Db 1050 GCTACATTGTGTGACCAAATGACTGGCATACTGGCAACAGATGTCAGTGCGGACGACGCG 1109
Qy 1084 CAAAAACTATTGGTTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAAT 1143
Db 1110 CAAAAACTGCTGGTTGGGCTCAACCAGCGCATAGTCGTCAACGGTCGCACCCAAAGAAAC 1169
Qy 1144 ACTAACACAATGCAGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCG 1203
Db 1170 ACCAATACCATGAAGAATTATCTTTTGCCCGTAGTGGCCCAGGCATTTGCTAGGTGGGCA 1229
Qy 1204 CGTGAACATCGTGCCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTT 1263
Db 1230 AAAGAATATAAGGAAGATCAAGAAGATGAGAGGCCACTAGGACTACGAGATAGACAGTTA 1289
Qy 1264 ACTATGGGCTGCTGCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCT 1323
Db 1290 GTTATGGGGTGCTGCTGGGCTTTTAGAAGGCACAAGATAACATCTATTTATAAGCGCCCA 1349
Qy 1324 GGTACGCAAACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTT 1383
Db 1350 GATACCCAAACCATCATCAAAGTGAACAGCGATTTCCACTCATTTGTGCTGCCCAGGATA 1409
Qy 1384 ACCAGCCACGGGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACT 1443
Db 1410 GGCAGTAACACATTGGAGATCGGGCTGAGAACAAGAATCAGGAAAATGCTAGAAGAGCGC 1469
Qy 1444 GTCAAACCCGCACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAA 1503
Db 1470 AAGGAGCCGTCACCTCTCATTACTGCCGAGGACGTACAAGAAGCTAAGTGCGCAGCCGAT 1529
Qy 1504 GAAGCTGAAGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCT 1563
Db 1530 GAGGCTAAGGAGGTGCGTGAAGCCGAGGAGTTGCGCGCTGCACTACCACCTTTGGCAGCA 1589
Qy 1564 GAAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGAGCA 1623
Db 1590 GATGTTGAGGAGCCCACTCTGGAAGCCGATGTTGACTTTATGCTACAAGAGGCTGGGGCC 1649
Qy 1624 GGTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAG 1683
Db 1650 GGCTCAGTGGAGACACCTCGTGGCTTGATAAAGGTTACCAGCTATGCCGGCGAGGACAAG 1709
Qy 1684 ATTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGT 1743
Db 1710 ATTGGCTCTTACGCAGTGCTTTCTCCACAGGCTGTACTCAAGAGTGAGAAACTATCTTGC 1769
Qy 1744 ATCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATAC 1803
Db 1770 ATTCACCCTCTCGCTGAACAAGTCATAGTGATAACACACTCTGGCCGAAAAGGGCGTTAT 1829
Qy 1804 AAAGTCGAGCCATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAA 1863
Db 1830 GCCGTGGAACCCTACCATGGAAAAGTAGTGGTGCCAGAGGGACATGCAATACCCGTCCAG 1889
Qy 1864 GACTTCCAGGCATTGAGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAAC 1923
Db 1890 GACTTTCAAGCTCTGAGTGAAAGTGCCACCATTGTGTACAACGAACGAGAGTTCGTAAAC 1949
Qy 1924 AGATACCTGCACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTAT 1983
Db 1950 AGGTATTTGCACCATATTGCCACACATGGAGGAGCGCTGAACACAGATGAAGAATATTAC 2009
Qy 1984 AAGACTGTAAAGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAG 2043
Db 2010 AAAATTGTCAAGCCCAGCGAGCACGACGGCGAATACCTGTATGACATCGACAGGAAACAG 2069
Qy 2044 TGTGTTAAGCGAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCA 2103
Db 2070 TGCGTCAAGAAAGAACTAGTCACTGGGCTAGGGCTTACAGGCGAGCTGGTGGATCCTCCC 2129
Qy 2104 TTTCACGAGTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCA 2163
Db 2130 TTCCATGAATTCGCCTACGAGAGCCTGAGAACACGACCGGCCGCTCCTTACCAAGTACCA 2189
Qy 2164 ACCATCGGAGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTG 2223
Db 2190 ACCATAGGGGTGTATGGCGTGCCGGGTTCAGGCAAGTCTGGCATCATTAAAAGCGCAGTC 2249
Qy 2224 ACTAAGAAAGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGAT 2283
Db 2250 ACCAAAAAAGATCTGGTGGTGAGCGCCAAGAAAGAGAACTGCGCAGAAATAATAAGGGAC 2309
Qy 2284 GTAAGGAGGATGAGACGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAAT 2343
Db 2310 GTCAAGAAAATGAAAGGGCTGGACGTCAATGCCAGAACTGTGGACTCAGTGCTCTTGAAT 2369
Qy 2344 GGGGTTAAGCACCCCGTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGG 2403
Db 2370 GGATGCAAACACCCTGTAGAGACCCTGTATATTGACGAAGCTTTTGCTTGTCATGCAGGC 2429
Qy 2404 ACGCTGCTGGCACTGATTGCCATCGTCAAACCTAAGAAAGTGGTATTGTGCGGGGACCCA 2463
Db 2430 ACTCTCAGAGCGCTCATAGCCATCATAAGACCTAAAAAGGCAGTGCTCTGCGGGGATCCA 2489
Qy 2464 AAACAATGCGGCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGC 2523
Db 2490 AAACAGTGCGGCTTTTTCAATATGATGTGCCTGAAAGTGCATTTTAACCACGAGATTTGC 2549
Qy 2524 ACTGAAGTGTACCATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATCGTC 2583
Db 2550 ACGCAGGTCTTCCACAAAAGCATCTCTCGTCGTTGCACTAAATCTGTGACTTCGGTCGTC 2609
Qy 2584 TCCACGCTCTTCTACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATCATC 2643
Db 2610 TCAACCTTGTTTTACGACAAAAGAATGAGAACGACGAACCCGAAAGAGACTAAGATTGAG 2669
Qy 2644 ATAGATACCACAGGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGA 2703
Db 2670 ATTGACACTACTGGCAGTACTAAACCAAAGCAGGACGATCTCATTCTCACTTGTTTCAGA 2729
Qy 2704 GGATGGGTGAAACAGCTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCTGCA 2763
Db 2730 GGGTGGGTAAAGCAGTTGCAAATAGATTACAAAGGCAACGAAATAATGACGGCAGCTGCC 2789
Qy 2764 TCGCAAGGACTTACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCA 2823
Db 2790 TCTCAAGGGCTGACCCGTAAAGGCGTGTATGCCGTTCGGTACAAGGTGAATGAAAATCCC 2849



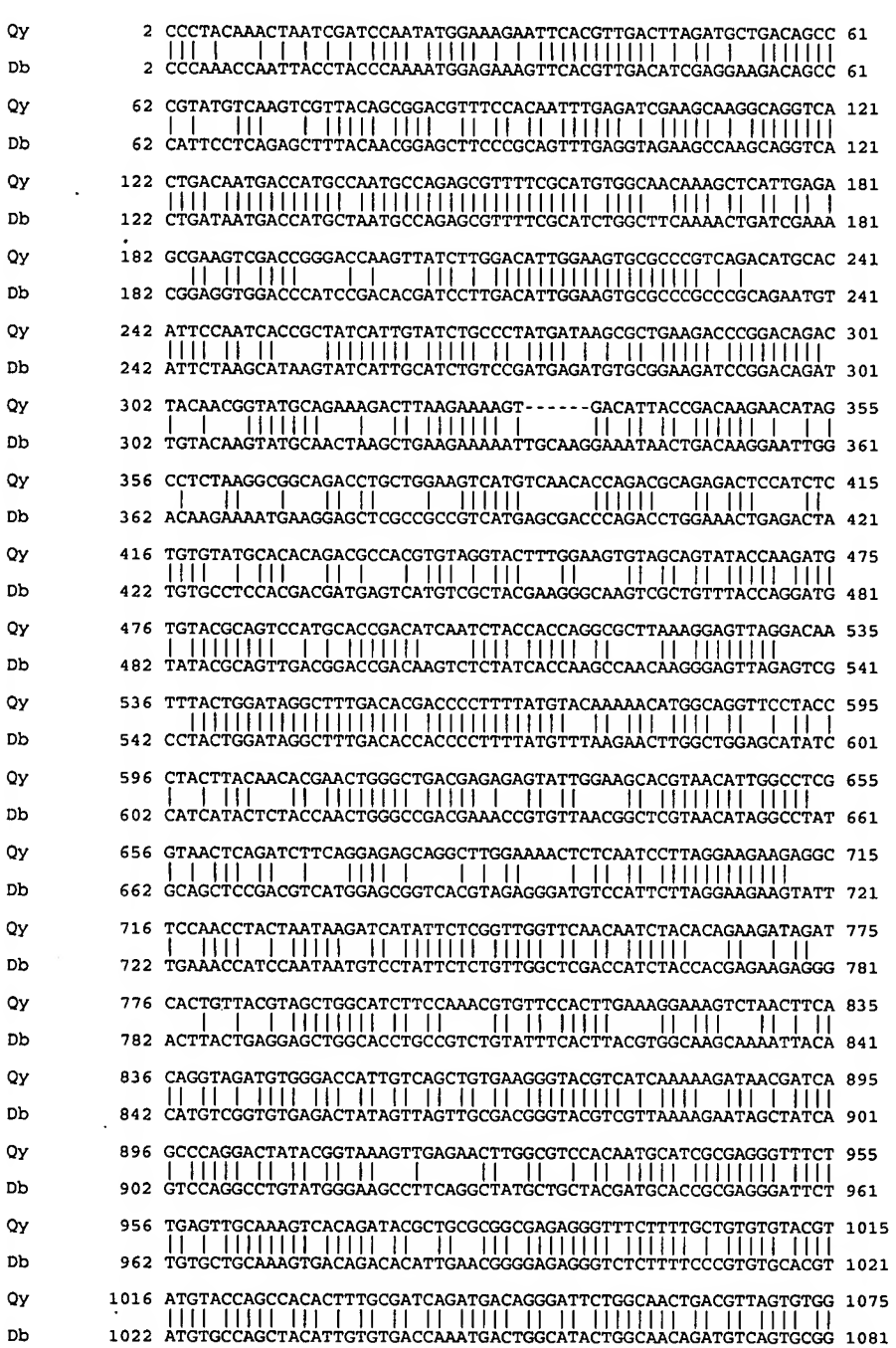
```
Db      5010 CGGAAGTAC 5018


RESULT 14
AF375051
LOCUS       AF375051               11423 bp    RNA     linear   VRL 26-JUN-2001
DEFINITION  Venezuelan equine encephalitis virus nonstructural polyprotein and
            structural polyprotein genes, complete cds.
ACCESSION   AF375051
VERSION     AF375051.1  GI:14549692
KEYWORDS    .
SOURCE      Venezuelan equine encephalitis virus
  ORGANISM  Venezuelan equine encephalitis virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
            Alphavirus; VEEV complex.
REFERENCE   1  (bases 1 to 11423)
  AUTHORS   Brault,A.C., Powers,A.M., Medina,G., Wang,E., Kang,W., Salas,R.A.,
            De Siger,J. and Weaver,S.C.
  TITLE     Potential sources of the 1995 Venezuelan equine encephalitis
            subtype IC epidemic
  JOURNAL   J. Virol. 75 (13), 5823-5832 (2001)
  MEDLINE   21286712
   PUBMED   11390583
REFERENCE   2  (bases 1 to 11423)
  AUTHORS   Weaver,S.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-2001) Pathology, Univ. Texas Medical Branch, 301
            University Blvd., Galveston, TX 77555-0609, USA
FEATURES             Location/Qualifiers
     source          1. .11423
                     /organism="Venezuelan equine encephalitis virus"
                     /mol_type="genomic RNA"
                     /strain="P676"
                     /db_xref="taxon:11036"
                     /note="isolated from mouse brain antigen"
     CDS             25. .7506
                     /note="nsP1-4"
                     /codon_start=1
                     /transl_except=(pos:5662. .5664,aa:OTHER)
                     /product="nonstructural polyprotein"
                     /protein_id="AAK66989.1"
                     /db_xref="GI:14549693"
                     /translation="MEKVHVDIEEDSPFLRALQRSFPQFEVEAKQVTDNDHANARAFS
                     HLASKLIETEVDPSDTILDIGSAPARRMYSKHKYHCICPMRCAEDPDRLYKYATKLKK
                     NCKEITDKELDKKMKELAAVMSDPDLETETMCLHDDESCRYEGQVAVYQDVYAVDGPT
                     SLYHQANKGVRVAYWIGFDTTPFMFKNLAGAYPSYSTNWADETVLTARNIGLCSSDVM
                     ERSRRGMSILRKKYLKPSNNVLFSVGSTIYHEKRDLLRSWHLPSVFHLRGKQNYTCRC
                     ETIVSCDGYVVKRIAISPGLYGKPSGYAATMHREGPLCCKVTDTLNGERVSFPVCTYV
                     PATLCDQMTGILATDVSADDAQKLLVGLNQRIVVNGRTQRNTNTMKNYLLPVVAQAFA
                     RWAKEYKEDQEDERPLGLRDRQLVMGCCWAFRRHKITSIYKRPDTQTIIKVNSDFHSF
                     VLPRIGSNTLEIGLRTRIRKMLEEHKEPSPLITAEDIQEAKCAADEAKEVREAEELRA
                     ALPPLAADFEEPTLEADVDLMLQEAGAGSVETPRGLIKVTSYAGEDKIGSYAVLSPQA
                     VLKSEKLSCIHPLAEQVIVITHSGRKGRYAVEPYHGKVVVPEGHAIPVQDFQALSESA
                     TIVYNEREFVNRYLHHIATHGGALNTDEEYYKTVKPSEHDGEYLYDIDRKQCVKKELV
                     TGLGLTGELVDPPFHEFAYESLRTRPAAPYQVPTIGVYGVPGSGKSGIIKSAVTKKDL
                     VVSAKKENCAEIIRDVKKMKGLDVNARTVDSVLLNGCKHPVETLYIDEAFACHAGTLR
                     ALIAIIRPKKAVLCGDPKQCGFFNMMCLKVHFNHEICTQVFHKSISRRCTKSVTSVVS
                     TLFYDKRMRTTNPKETKIEIDTTGSTKPKQDDLILTCFRGWVKQLQIDYKGNEIMTAA
                     ASQGLTRKGVYAVRYKVNENPLYAPTSEHVNVLLTRTEDRIVWKTLAGDPWIKTLTAK
                     YPGNFTATIEEWQAEHDAIMRHILERPDPTDVFQNKANVCWAKALVPVLKTAGIDMTT
                     EQWNTVDYFETDKAHSAEIVLNQLCVRFFGLDLDSGLFSAPTVPLSIRNNHWDNSPSP
                     NMYGLNKEVVRQLSRRYPQLPRAVATGRVYDMNTGTLRNYDPRINLVPVNRRLPHALV
                     LHHNEHPQSDFSSFVSKLKGRTVLVVGEKLSVPGKKVDWLSDQPEATFRARLDLGIPG
                     DVPKYDIVFINVRTPYKYHHYQQCEDHAIKLSMLTKKACLHLNPGGTCVSIGYGYADR
                     ASESIIGAIARQFKFSRVCKPKSSHEETEVLFVFIGYDRKARTHNPYKLSSTLTNIYT
                     GSRLHEAGCAPSYHVVRGDIATATEGVIINAANSKGQPGGGVCGALYKKFPESFDLQP
                     IEVGKARLVKGVTKHIIHAVGPNFNKVSEVEGDKQLAEAYESIAKIVNDNNYKSVAIP
                     LLSTGIFSGNKDRLTQSLNHLLTALDTTDADVAIYCRDKKWEMTLKEAVARREAVEEI
                     CISDDSSVTEPDAELVRVHPKSSLAGRKGYSTSDGKTFSYLEGTKFHQAAKDIAEINA
                     MWPVATEANEQVCMYILGESMSSIRSKCPVEESEASTPPSTLPCLCIHAMTPERVQRL
                     KASRPEQITVCSSFPLPKYRITGVQKIQCSQPILFSPKVPAYIHPRKYLVETPPVEET
                     PESPAENQSTEGTPEQPALVNVDATRTRMPEPIIIEEEEEDSISLLSDGPTHQVLQVE
                     ADIHGSPSVSSSSWSIPHASDFDVDSLSILDTLDGASVTSEAVSAETNSYFARSMEFR
                     ARPVPAPRTVFRNPPHPAPRTRTPPLAHSRASSRTSLVSTPPGVNRVITREELEALTP
                     SRAPSRSASRTSLVSNPPGVNRVITREEFEAFVAQQQXRFDAGAYIFSSDTGQGHLQQ
                     KSVRQTVLSEVVLERTELEISYAPRLDQEKEELLRKKLQLNPTPANRSRYQSRRVENM
                     KAITARRILQGLGHYLKAEGKVECYRTLHPVPLYSSSVNRAFSSPKVAVEACNAMLKE
                     NFPTVASYCIIPEYDAYLDMVDGASCCLDTASFCPAKLRSFPKKHSYLEPTIRSAVPS
                     AIQNTLQNVLAAATKRNCNVTQMRELPVLDSAAFNVECFKKYACNNEYWETFKENPIR
                     LTEENVVNYITKLKGPKAAALFAKTHNLNMLQDIPMDRFVMDLKRDVKVTPGTKHTEE
                     RPKVQVIQAADPLATAYLCGIHRELVRRLNAVLLPNIHTLFDMSAEDFDAIIAEHFQP
                     GDCVLETDIASFDKSEDDAMALTALMILEDLGVDAELLTLIEAAFGEISSIHLPTKTK
                     FKFGAMMKSGMFLTLFVNTVINIVIASRVLRERLTGSPCAAFIGDDNIVKGVKSDKLM
                     ADRCATWLNMEVKIIDAVVGEKAPYFCGGFILCDSVTGTACRVADPLKRLFKLGKPLA
                     VDDEHDDDRRRALHEESTRWNRVGILPELCKAVESRYETVGTSIIVMAMTTLASSVKS
                     FSYLRGAPITLYG"
     CDS             7542. .11309
                     /note="encodes capsid, E2, and E1 envelope glycoproteins"
                     /codon_start=1
                     /product="structural polyprotein"
                     /protein_id="AAK66990.1"
                     /db_xref="GI:14549694"
                     /translation="MFPFQPMYPMQPMPYRNPFAAPRRPWFPRTDPFLAMQVQELTRS
                     MANLTFKQRRDAPPEGPPAKKPKREAPQKQKGGGQGKKKKNQGKKKAKTGPPNPKAQS
                     GNKKKPNKKPGKRQRMVMKLESDKTFPIMLEGKINGYACVVGGKLFRPMHVEGKIDND
                     VLAALKTKKASKYDLEYADVPQNMRADTFKYTHEKPQGYYSWHHGAVQYENGRFTVPK
                     GVGAKGDSGRPILDNQGRVVAIVLGGVNEGSRTALSVVMWNEKGVTVKYTPENCEQWS
                     LVTTMCLLANVTFPCAEPPICYDRKPAETLAMLSVNVDNPGYDELLEAAVKCPGRKRR
                     STEELFKEYKLTRPYMARCIRCAVGSCHSPIAIEAVKSDGHDGYVRLQTSSQYGLDSS
                     GNLKGRTMRYDMHGTIEEIPLHQVSLHTSRPCHIVDGHGYFLLARCPAGDSITMEFKK
                     GSVTHSCSVPYEVKFNPVGRELYTHPPEHGAEQACQVYAHDAQNRGAYVEMHLPGSEV
                     DSSLISLSGSSVTVTPPVGTSALVECKCGGTKISETINKAKQFSQCTKKEQCRAYRLQ
                     NDKWVYNSDKLPKAAGATLKGKLHVPFLLADGKCTVPLAPEPMITFGFRSVSLKLHPK
                     NPTYLTTRQLADEPHYTHELISEPAVRNFTVTEKGWEFVWGNHPPKRFWAQETAPGNP
                     HGLPHEVITHYYHRYPMSTILGLSICAAIVTVSVAASTWLFCKSRVSCLTPYRLTPNA
                     RMPLCLAVLCCARTARAETTWESLDHLWNNNQQMFWIQLLIPLAALIVVTRLLKCVCC
                     VVPFLVVAGAAGAGAYEHATTMPSQAGISYNTIVNRAGYAPLPISITPTKIKLIPTVN
                     LEYVTCHYKTGMDSPAIKCCGSQECTPTNRPDEQCKVFTGVYPFMWGGAYCFCDTENT
                     QVSKAYVMKSDDCLADHAEAYKAHTASVQAFLNITVGEHSIVTTVYVNGETPVNFNGV
                     KLTAGPLSTAWTPFDRKIVQYAGEIYNYDFPEYGAGQPGAFGDIQSRTVSSSDLYANT
                     NLVLQRPKAGAIHVPYTQAPSGFEQWKKDKAPSLKFTAPFGCEIYTNPIRAENCAVGS
                     IPLAFDIPDALFTRVSETPTLSAAECTLNECVYSSDFGGIATVKYSASKSGKCAVHVP
                     SGTATLKEAAVELTEQGSATIHFSTANIHPEFRLQICTSYVTCKGDCHPPKDHIVTHP
                     QYHAQTFTAAVSKTAWTWLTSLLGGSAVIIIGLVLATIVAMYVLTNQKHN"
BASE COUNT     3227 a   2819 c   2902 g   2475 t
ORIGIN

  Query Match           42.4%;  Score 2119;  DB 14;  Length 11423;
  Best Local Similarity   64.5%;  Pred. No. 0;
  Matches 3223;  Conservative    0;  Mismatches 1750;  Indels   24;  Gaps    3;

Qy          2 CCCTACAAACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCC 61
              ||| |   | | | | | |||| |||||  | | |||||||||||| | || |  |||||||
Db          2 CCCAAACCAATTACCTACCCAAAATGGAGAAAGTTCACGTTGACATCGAGGAAGACAGCC 61

Qy         62 CGTATGTCAAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCA 121
              | |   |||   | ||||| ||||  || || || |||||| | |||||  | ||||||||
Db         62 CATTCCTCAGAGCTTTACAACGGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCA 121

Qy        122 CTGACAATGACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGA 181
              |||| |||||||||||| |||||||||||||||||||||| ||||  |||| || || || |
Db        122 CTGATAATGACCATGCTAATGCCAGAGCGTTTTCGCATCTGGCTTCAAAACTGATCGAAA 181

Qy        182 GCGAAGTCGACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCAC 241
                || || ||||     | |    ||| | |||||||||||||||||| | |
Db        182 CGGAGGTGGACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCCGCCCGCAGAATGT 241

Qy        242 ATTCCAATCACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGAC 301
              |||| || ||    |||||||| ||||| || |||| | | || ||||| |||||||||
Db        242 ATTCTAAGCATAAGTATCATTGCATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGAT 301

Qy        302 TACAACGGTATGCAGAAAGACTTAAGAAAAGT------GACATTACCGACAAGAACATAG 355
              |  |   |||||||   |  || |||||||  |      || || || ||||||  |  | |
Db        302 TGTACAAGTATGCAACTAAGCTGAAGAAAAATTGCAAGGAAATAACTGACAAGGAATTGG 361

Qy        356 CCTCTAAGGCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTC 415
```

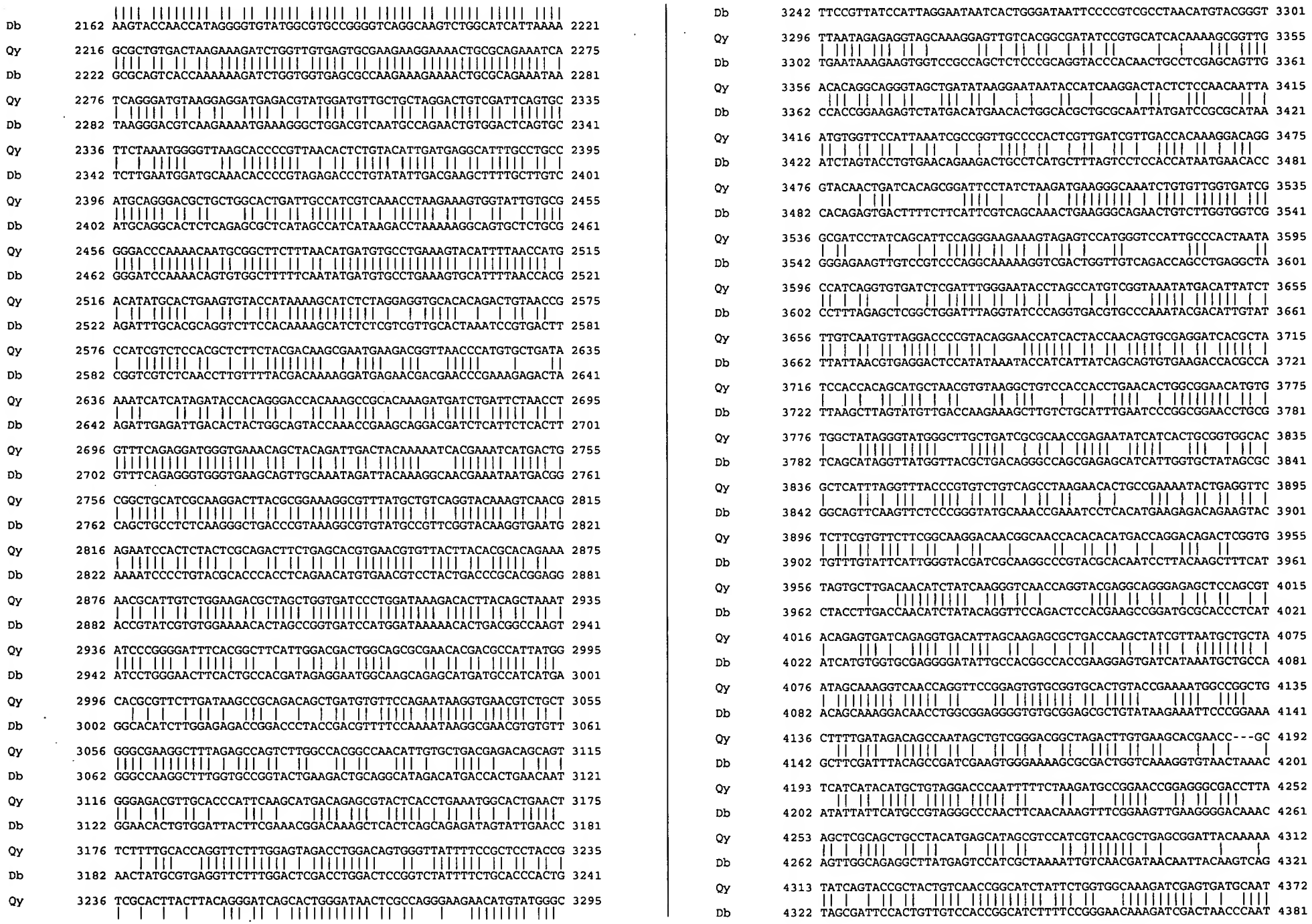

Db 362 ACAAGAAAATGAAGGAGCTCGCCGCCGTCATGAGCGACCCAGACCTGGAAACTGAGACTA 421
Qy 416 TGTGTATGCACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATG 475
Db 422 TGTGCCTCCACGACGATGAGTCATGTCGCTACGAGGGGCAAGTCGCTGTTTACCAGGATG 481
Qy 476 TGTACGCAGTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAA 535
Db 482 TATACGCAGTTGACGGACCGACAAGTCTCTATCACCAAGCCAACAAGGGAGTTAGAGTCG 541
Qy 536 TTTACTGGATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACC 595
Db 542 CCTACTGGATAGGCTTTGACACCACCCCTTTTATGTTTAAGAACTTGGCTGGAGCATATC 601
Qy 596 CTACTTACAACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCG 655
Db 602 CATCATACTCTACCAACTGGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCCTAT 661
Qy 656 GTAACTCAGATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGC 715
Db 662 GCAGCTCCGACGTCATGGAGCGGTCACGTAGAGGGATGTCCATTCTTAGGAAGAAGTATT 721
Qy 716 TCCAACCTACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGAT 775
Db 722 TGAAACCATCCAATAATGTCCTATTCTCTGTTGGCTCGACCATCTACCACGAGAAGAGGG 781
Qy 776 CACTGTTACGTAGCTGGCATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCA 835
Db 782 ACTTACTGAGGAGCTGGCACCTGCCGTCTGTATTTCACTTACGTGGCAAGCAAAATTACA 841
Qy 836 CAGGTAGATGTGGGACCATTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCA 895
Db 842 CATGTCGGTGTGAGACTATAGTTAGTTGCGACGGGTACGTCGTTAAAAGAATAGCTATCA 901
Qy 896 GCCCAGGACTATACGGTAAAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCT 955
Db 902 GTCCAGGCCTGTATGGGAAGCCTTCAGGCTATGCTGCTACGATGCACCGCGAGGGATTCT 961
Qy 956 TGAGTTGCAAAGTCACAGATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGT 1015
Db 962 TGTGCTGCAAAGTGACAGACACATTGAACGGGGAGAGGGTCTCTTTTCCCGTGTGCACGT 1021
Qy 1016 ATGTACCAGCCACACTTTGCGATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGG 1075
Db 1022 ATGTGCCAGCTACATTGTGTGACCAAATGACTGGCATACTGGCAACAGATGTCAGTGCGG 1081
Qy 1076 ATGACGCACAAAAACTATTGGTTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGC 1135
Db 1082 ACGACGCGCAAAAACTGCTGGTTGGGCTCAACCAGCGCATAGTCGTCAACGGTCGCACCC 1141
Qy 1136 AAAGAAATACTAACACAATGCAGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCA 1195
Db 1142 AAAGAAACACCAATACCATGAAGAATTATCTTTTGCCCGTAGTGGCCCAGGCATTTGCTA 1201
Qy 1196 GGTGGGCGCGTGAACATCGTGCCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGC 1255
Db 1202 GGTGGGCAAAGGAATATAAGGAAGATCAAGAAGATGAGAGGCCACTAGGACTACGAGATA 1261
Qy 1256 GCACTCTTACTATGGGCTGCTGCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACA 1315
Db 1262 GACAGTTAGTCATGGGGTGCTGCTGGGCTTTTAGAAGGCACAAGATAACATCTATTTATA 1321
Qy 1316 AGAAGCCTGGTACGCAAACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTC 1375
Db 1322 AGCGCCCAGATACCCAAACCATCATCAAAGTGAACAGCGATTTCCACTCATTCGTGCTGC 1381
Qy 1376 CACGCCTTACCAGCCACGGGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTG 1435
Db 1382 CCAGGATAGGCAGTAACACACTGGAGATCGGGCTGAGAACGAGAATCAGGAAAATGCTAG 1441
Qy 1436 AACCAACTGTCAAACCCGCACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCT 1495
Db 1442 AAGAGCACAAGGAGCCGTCACCTCTCATTACTGCCGAGGACATACAAGAGGCTAAGTGCG 1501
Qy 1496 TACAGCAAGAAGCTGAAGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCT 1555
Db 1502 CAGCCGATGAGGCTAAGGAAGTGCGTGAAGCCGAGGAGCTGCGCGCTGCACTACCACCTT 1561
Qy 1556 TGCTCCCTGAAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGG 1615
Db 1562 TGGCAGCTGATTTTGAGGAGCCCACTCTGGAAGCCGATGTCGACTTGATGTTACAAGAGG 1621
Qy 1616 CAGGAGCAGGTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCG 1675
Db 1622 CTGGGGCCGGCTCAGTGGAGACACCTCGTGGCTTGATAAAGGTTACCAGCTATGCCGGCG 1681
Qy 1676 AAGAGAAGATTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAAC 1735
Db 1682 AGGACAAGATCGGCTCTTACGCAGTGCTTTCTCCACAGGCTGTACTCAAGAGTGAGAAAC 1741
Qy 1736 TGGCGTGTATCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAG 1795
Db 1742 TATCTTGCATTCACCCTCTCGCTGAACAAGTCATAGTGATAACACACTCTGGCCGAAAAG 1801
Qy 1796 GGAGATACAAAGTCGAGCCATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCC 1855
Db 1802 GGCGTTATGCCGTGGAACCCTACCATGGAAAAGTAGTGGTGCCAGAGGGACATGCAATAC 1861
Qy 1856 CTGTTCAAGACTTCCAGGCATTGAGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGT 1915
Db 1862 CCGTCCAGGACTTTCAAGCTCTGAGTGAAAGTGCCACCATCGTGTACAACGAACGAGAGT 1921
Qy 1916 TCGTAAACAGATACCTGCACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAG 1975
Db 1922 TCGTAAACAGGTACCTGCACCATATTGCCACACATGGAGGAGCGCTGAACACAGATGAAG 1981
Qy 1976 AGTACTATAAGACTGTAAAGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGACG 2035
Db 1982 AATATTACAAAACTGTCAAGCCCAGCGAGCACGACGGCGAATACCTGTATGACATCGACA 2041
Qy 2036 CACGAAAGTGTGTTAAGCGAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAG 2095
Db 2042 GGAAACAATGCGTCAAGAAAGAATTAGTCACTGGGCTAGGGCTTACAGGCGAGCTGGTGG 2101
Qy 2096 ATCCACCATTTCACGAGTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACA 2155
Db 2102 ATCCTCCCTTCCATGAATTTGCCTACGAGAGTCTGAGAACACGTCCGGCCGCTCCTTACC 2161
Qy 2156 AAGTCCCAACCATCGGAGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAA 2215
Db 2162 AAGTACCAACCATAGGGGTGTATGGCGTGCCGGGGTCAGGCAAGTCTGGCATCATTAAAA 2221
Qy 2216 GCGCTGTGACTAAGAAAGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCA 2275
Db 2222 GCGCAGTCACCAAAAAAGATCTGGTGGTGAGCGCCAAGAAAGAAAACTGCGCAGAAATAA 2281
Qy 2276 TCAGGGATGTAAGGAGGATGAGACGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGC 2335
Db 2282 TAAGGGACGTCAAGAAAATGAAAGGGCTGGACGTCAATGCCAGAACTGTGGACTCAGTGC 2341
Qy 2336 TTCTAAATGGGGTTAAGCACCCCGTTAACACTCTGTACATTGATGAGGCATTTGCCTGCC 2395
Db 2342 TCTTGAATGGATGCAAACACCCCGTAGAGACCCTGTATATTGACGAAGCTTTTGCTTGTC 2401
Qy 2396 ATGCAGGGACGCTGCTGGCACTGATTGCCATCGTCAAACCTAAGAAAGTGGTATTGTGCG 2455
Db 2402 ATGCAGGCACTCTCAGAGCGCTCATAGCCATCATAAGACCTAAAAAGGCAGTGCTCTGCG 2461
Qy 2456 GGGACCCAAAACAATGCGGCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATG 2515
Db 2462 GGGATCCAAAACAGTGTGGCTTTTTCAATATGATGTGCCTGAAAGTGCATTTTAACCACG 2521
Qy 2516 ACATATGCACTGAAGTGTACCATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCG 2575
Db 2522 AGATTTGCACGCAGGTCTTCCACAAAAGCATCTCTCGTCGTTGCACTAAATCCGTGACTT 2581



```
Db       4742 TGAGCAGCATTAGGTCGAAATGCCCCGTCGAGGAGTCGGAAGCCTCCACACCACCTAGCA 4801

Qy       4778 CACTTCCATGCCTGTGTAATTACGCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTG 4837

Db       4802 CGCTGCCTTGCTTGTGCATCCATGCTATGACTCCAGAAAGAGTACAACGCCTAAAAGCCT 4861

Qy       4838 CGAAGAAAGAACAGTTCGCCGTATGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAG 4897

Db       4862 CACGTCCAGAACAAATTACTGTGTGCTCATCCTTTCCATTGCCGAAGTATAGAATCACTG 4921

Qy       4898 GCGTGCAGAAGCTACAGTGCAGCAAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTG 4957

Db       4922 GTGTGCAGAAGATCCAGTGCTCCCAGCCTATACTGTTCTCACCGAAGGTGCCTGCGTACA 4981

Qy       4958 TACACCCCAGGAAGTAC 4974

Db       4982 TTCATCCACGGAAGTAC 4998
```

```
RESULT 15
VEU55342
LOCUS       VEU55342               11423 bp    RNA     linear   VRL 15-JAN-2002
DEFINITION  Venezuelan equine encephalitis virus nonstructural polyprotein and
            structural polyprotein genes, complete cds.
ACCESSION   U55342
VERSION     U55342.2  GI:18152933
KEYWORDS    .
SOURCE      Venezuelan equine encephalitis virus
  ORGANISM  Venezuelan equine encephalitis virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
            Alphavirus; VEEV complex.
REFERENCE   1  (bases 1 to 11423)
  AUTHORS   Weaver,S.C., Salas,R., Rico-Hesse,R., Ludwig,G.V., Oberste,M.S.,
            Boshell,J. and Tesh,R.B.
  TITLE     Re-emergence of epidemic Venezuelan equine encephalomyelitis in
            South America. VEE Study Group
  JOURNAL   Lancet 348 (9025), 436-440 (1996)
  MEDLINE   96332169
   PUBMED   8709783
REFERENCE   2  (bases 1 to 11423)
  AUTHORS   Weaver,S.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-1996) Scott C. Weaver, Pathology, Univ. TX Med.
            Branch, 301 University Ave., Galveston, TX 77555-0609, USA
REFERENCE   3  (bases 1 to 11423)
  AUTHORS   Weaver,S.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2001) Scott C. Weaver, Pathology, Univ. TX Med.
            Branch, 301 University Ave., Galveston, TX 77555-0609, USA
  REMARK    Sequence update by submitter
COMMENT     On Jan 15, 2002 this sequence version replaced gi:2335113.
FEATURES             Location/Qualifiers
     source          1. .11423
                     /organism="Venezuelan equine encephalitis virus"
                     /mol_type="genomic RNA"
                     /strain="V198"
                     /variety="antigenic IC"
                     /db_xref="taxon:11036"
     CDS             25. .7506
                     /note="contains low efficiency read through at stop codon;
                     encodes nsP1, nsP2, nsP3 and nsP4 proteins"
                     /codon_start=1
                     /transl_except=(pos:5662. .5664,aa:OTHER)
                     /product="nonstructural polyprotein"
                     /protein_id="AAL61965.1"
                     /db_xref="GI:18152935"
                     /translation="MEKVHVDIEEDSPFLRALQRSFPQFEVEAKQVTDNDHANARAFS
                     HLASKLIETEVDPSDTILDIGSAPARRMYSKHKYHCICPMRCAEDPDRLYKYATKLKK
                     NCKEITDKELDKKMKELAAVMSDPDLETETMCLHDDESCRYEGQVAVYQDVYAVDGPT
                     SLYHQANKGVRVAYWIGFDTTPFMFKNLAGAYPSYSTNWADETVLTARNIGLCSSDVM
                     ERSRRGMSILRKKYLKPSNNVLFSVGSTIYHEKRDLLRSWHLPSVFHLRGKQNYTCRC
                     ETIVSCDGYVVKRIAISPGLYGKPSGYAATMHREGFLCCKVTDTLNGERVSFPVCTYV
                     PATLCDQMTGILATDVSADDAQKLLVGLNQRIVVNGRTQRNTNTMKNYLLPVVAQAFA
                     RWAKEYKEDQEDERPLGLRDRQLVMGCCWAFRRHKITSIYKRPDTQTIIKVNSDFHSF
                     VLPRIGSNTLEIGLRTRIRKMLEEHKEPSPLITAEDIQEAKCAADEAKEVREAEELRA
                     ALPPLAADFEEPTLEADVDLMLQEAGAGSVETPRGLIKVTSYAGEDKIGSYAVLSPQA
                     VLKSEKLSCIHPLAEQVIVITHSGRKGRYAVEPYHGKVVVPEGHAIPVQDFQALSESA
                     TIVYNEREFVNRYLHHIATHGGALNTDEEYYKTVKPSEHDGEYLYDIDRKQCVKKELV
                     TGLGLTGELVDPPFHEFAYESLRTRPAAPYQVPTIGVYGVPGSGKSGIIKSAVTKKDL
                     VVSAKKENCAEIIRDVKKMKGLDVNARTVDSVLLNGCKHPVETLYIDEAFACHAGTLR
                     ALIAIIRPKKAVLCGDPKQCGFFNMMCLKVHFNHEICTQVFHKSISRRCTKSVTSVVS
                     TLFYDKRMRTTNPKETKIEIDTTGSTKPKQDDLILTCFRGWVKQLQIDYKGNEIMTAA
                     ASQGLTRKGVYAVRYKVNENPLYAPTSEHVNVLLTRTEDRIVWKTLAGDPWIKTLTAK
                     YPGNFTATIEEWQAEHDAIMRHILERPDPTDVFQNKANVCWAKALVPVLKTAGIDMTT
                     EQWNTVDYFETDKAHSAEIVLNQLCVRFFGLDLDSGLFSAPTVPLSIRNNHWDNSPSP
                     NMYGLNKEVVRQLSRRYPQLPRAVATGRVYDMNTGTLRNYDPRINLVPVNRRLPHALV
                     LHHNEHPQSDFSSFVSKLKGRTVLVVGEKLSVPGKKVDWLSDQPEATFRARLDLGIPG
                     DVPKYDIVFINVRTPYKYHHYQQCEDHAIKLSMLTKKACLHLNPGGTCVSIGYGYADR
                     ASESIIGAIARQFKFSRVCKPKSSHEETEVLFVFIGYDRKARTHNPYKLSSTLTNIYT
                     GSRLHEAGCAPSYHVVRGDIATATEGVIINAANSKGQPGGGVCGALYKKFPESFDLQP
                     IEVGKARLVKGVTKHIIHAVGPNFNKVSEVEGDKQLAEAYESIAKIVNDNNYKSVAIP
                     LLSTGIFSGNKDRLTQSLNHLLTALDTTDADVAIYCRDKKWEMTLKEAVARREAVEEI
                     CISDDSSVTEPDAELVRVHPKSSLAGRKGYSTSDGKTFSYLEGTKFHQAAKDIAEINA
                     MWPVATEANEQVCMYILGESMSSIRSKCPVEESEASTPPSTLPCLCIHAMTPERVQRL
                     KASRPEQITVCSSFPLPKYRITGVQKIQCSQPILFSPKVPAYIHPRKYLVETPPVEET
                     PESPAENQSTEGTPEQPALVNVDATRTRMPEPIIIEEEEEDSISLLSDGPTHQVLQVE
                     ADIHGSPSVSSSSWSIPHASDFDVDSLSILDTLDGASVTSEAASAETNSYFARSMEFR
                     ARPVPAPRTVFRNPPHPAPRTRTPSLAHSRASSRTSLVSTPPGVNRVITREELEALTP
                     SRAPSRSASRTSLVSNPPGVNRVITREEFEAFVAQQQXRFDAGAYIFSSDTGQGHLQQ
                     KSVRQTVLSEVVLERTELEISYAPRLDQEKEELLRKKLQLNPTPANRSRYQSRRVENM
                     KAITARRILQQLGHYLKAEGKVECYRTLHPVPLYSSSVNRAFSSPKVAVEACNAMLKE
                     NFPTVASYCIIPEYDAYLDMVDGASCCLDTASFCPAKLRSFPKKHSYLEPTIRSAVPS
                     AIQNTLQNVLAAATKRNCNVTQMRELPVLDSAAFNVECFKKYACNNEYWETFKENPIR
                     LTEENVVNYITKLKGPKAAALFAKTHNLNMLQDIPMDRFVMDLKRDVKVTPGTKHTEE
                     RPKVQVIQAADPLATAYLCGIHRELVRRLNAVLLPNIHTLFDMSAEDFDAIIAEHFQP
                     GDCVLETDIASFDKSEDDAMALTALMILEDLGVDAELLTLIEAAFGEISSIHLPTKTK
                     FKFGAMMKSGMFLTLFVNTVINIVIASRVLRERLTGSPCAAFIGDDNIVKGVKSDKLM
                     ADRCATWLNMEVKIIDAVVGEKAPYFCGGFILCDSVTGTACRVADPLKRLFKLGKPLA
                     ADDEHDDDRRRALHEESTRWNRVGILPELCKAVESRYETVGTSIIVMAMTTLASSVKS
                     FSYLRGAPITLYG"
     CDS             7542. .11309
                     /note="encodes capsid protein and E2 and E1 envelope
                     glycoproteins"
                     /codon_start=1
                     /product="structural polyprotein"
                     /protein_id="AAC58460.2"
                     /db_xref="GI:18152934"
                     /translation="MFPFQPMYPMQPMPYRNPFAAPRRPWFPRTDPFLAMQVQELTRS
                     MANLTFKQRRDAPPEGPPAKKPKREAPQKQKGGGQGKKKKNQGKKKAKTGPPNPKAQS
                     GNKKKPNKKPGKRQRMVMKLESDKTFPIMLEGKINGYACVVGGKLFRPMHVEGKIDND
                     VLAALKTKKASKYDLEYADVPQNMRADTFKYTHEKPQGYYSWHHGAVQYENGRFTVPK
                     GVGAKGDSGRPILDNQGRVVAIVLGGVNEGSRTALSVVMWNEKGVTVKYTPENCEQWS
                     LVTTMCLLANVTFPCAEPPICYDRKPAETLAMLSVNVDNPGYDELLEAAVKCPGRKRR
                     STEELFKEYKLTRPYMARCIRCAVGSCHSPIAIEAVKSDGHDGYVRLQTSSQYGLDSS
                     GNLKGRTMRYDMHGTIEEIPLHQVSLHTSRPCHIVDGHGYFLLARCPAGDSITMEFKK
                     GSVTHSCSVPYEVKFNPVGRELYTHPPEHGAEQACQVYAHDAQNRGAYVEMHLPGSEV
                     DSSLISLSGSSVTVTPPVGTSALVECKCGGTKISETINKAKQFSQCTKKEQCRAYRLQ
                     NDKWVYNSDKLPKAAGATLKGKLHVPFLLADGKCTVPLAPEPMITFGFRSVSLKLHPK
                     NPTYLTTRQLADEPHYTHELISEPAVRNFTVTEKGWEFVWGNHPPKRFWAQETAPGNP
                     HGLPHEVITHYYHRYPMSTILGLSICAAIVTVSVAASTWLFCKSRVSCLTPYRLTPNA
                     RMPLCLAVLCCARTARAETTWESLDHLWNNNQQMFWIQLLIPLAALIVVTRLLKCVCC
                     VVPFLVVAGAAGAGAYEHATTMPSQAGISYNTIVNRAGYAPLPISITPTKIKLIPTVN
                     LEYVTCHYKTGMDSPAIKCCGSQECTPTNRPDEQCKVFTGVYPFMWGGAYCFCDTENT
                     QVSKAYVMKSDDCLADHAEAYKAHTASVQAFLNITVGEHSIVTTVYVNGETPVNFNGV
                     KLTAGPLSTAWTPFDRKIVQYAGEIYNYDFPEYGAGQPGAFGDIQSRTVSSSDLYANT
                     NLVLQRPKAGAIHVPYTQAPSGFEQWKKDKAPSLKFTAPFGCEIYTNPIRAENCAVGS
                     IPLAFDIPDALFTRVSETPTLSAAECTLNECVYSSDFGGIATVKYSASKSGKCAVHVP
                     SGTATLKEAAVELTEQGSATIHFSTANIHPEFRLQICTSYVTCKGDCHPPKDHIVTHP
                     QYHAQTFTAAVSKTAWTWLTSLLGGSAVIIIIGLVLATIVAMYVLTNQKHN"
BASE COUNT     3230 a   2820 c   2900 g   2473 t
ORIGIN

  Query Match           42.4%;  Score 2119;  DB 14;  Length 11423;
  Best Local Similarity   64.5%;  Pred. No. 0;
  Matches 3223;  Conservative    0;  Mismatches 1750;  Indels   24;  Gaps    3;
```


Qy 2 CCCTACAAACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCC 61
Db 2 CCCAAACCAATTACCTACCCAAAATGGAGAAAGTTCACGTTGACATCGAGGAAGACAGCC 61
Qy 62 CGTATGTCAAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCA 121
Db 62 CATTCCTCAGAGCTTTACAACGGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCA 121
Qy 122 CTGACAATGACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGA 181
Db 122 CTGATAATGACCATGCTAATGCCAGAGCGTTTTCGCATCTGGCTTCAAAACTGATCGAAA 181
Qy 182 GCGAAGTCGACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCAC 241
Db 182 CGGAGGTGGACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCCGCCCGCAGAATGT 241
Qy 242 ATTCCAATCACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGAC 301
Db 242 ATTCTAAGCATAAGTATCATTGCATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGAT 301
Qy 302 TACAACGGTATGCAGAAAGACTTAAGAAAAGT------GACATTACCGACAAGAACATAG 355
Db 302 TGTACAAGTATGCAACTAAGCTGAAGAAAAATTGCAAGGAAATAACTGACAAGGAATTGG 361
Qy 356 CCTCTAAGGCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTC 415
Db 362 ACAAGAAAATGAAGGAGCTCGCCGCCGTCATGAGCGACCCAGACCTGGAAACTGAGACTA 421
Qy 416 TGTGTATGCACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATG 475
Db 422 TGTGCCTCCACGACGATGAGTCATGTCGCTACGAAGGGCAAGTCGCTGTTTACCAGGATG 481
Qy 476 TGTACGCAGTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAA 535
Db 482 TATACGCAGTTGACGGACCGACAAGTCTCTATCACCAAGCCAACAAGGGAGTTAGAGTCG 541
Qy 536 TTTACTGGATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACC 595
Db 542 CCTACTGGATAGGCTTTGACACCACCCCTTTTATGTTTAAGAACTTGGCTGGAGCATATC 601
Qy 596 CTACTTACAACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCG 655
Db 602 CATCATACTCTACCAACTGGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCCTAT 661
Qy 656 GTAACTCAGATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGC 715
Db 662 GCAGCTCCGACGTCATGGAGCGGTCACGTAGAGGGATGTCCATTCTTAGGAAGAAGTATT 721
Qy 716 TCCAACCTACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGAT 775
Db 722 TGAAACCATCCAATAATGTCCTATTCTCTGTTGGCTCGACCATCTACCACGAGAAGAGGG 781
Qy 776 CACTGTTACGTAGCTGGCATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCA 835
Db 782 ACTTACTGAGGAGCTGGCACCTGCCGTCTGTATTTCACTTACGTGGCAAGCAAAATTACA 841
Qy 836 CAGGTAGATGTGGGACCATTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCA 895
Db 842 CATGTCGGTGTGAGACTATAGTTAGTTGCGACGGGTACGTCGTTAAAAGAATAGCTATCA 901
Qy 896 GCCCAGGACTATACGGTAAAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCT 955
Db 902 GTCCAGGCCTGTATGGGAAGCCTTCAGGCTATGCTGCTACGATGCACCGCGAGGGATTCT 961
Qy 956 TGAGTTGCAAAGTCACAGATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGT 1015
Db 962 TGTGCTGCAAAGTGACAGACACATTGAACGGGGAGAGGGTCTCTTTTCCCGTGTGCACGT 1021
Qy 1016 ATGTACCAGCCACACTTTGCGATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGG 1075
Db 1022 ATGTGCCAGCTACATTGTGTGACCAAATGACTGGCATACTGGCAACAGATGTCAGTGCGG 1081
Qy 1076 ATGACGCACAAAAACTATTGGTTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGC 1135
Db 1082 ACGACGCGCAAAAACTGCTGGTTGGGCTCAACCAGCGCATAGTCGTCAACGGTCGCACCC 1141
Qy 1136 AAAGAAATACTAACACAATGCAGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCA 1195
Db 1142 AAAGAAACACCAATACCATGAAGAATTATCTTTTGCCCGTAGTGGCCCAGGCATTTGCTA 1201
Qy 1196 GGTGGGCGCGTGAACATCGTGCCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGC 1255
Db 1202 GGTGGGCAAAGGAATATAAGGAAGATCAAGAAGATGAGAGGCCACTAGGACTACGAGATA 1261
Qy 1256 GCACTCTTACTATGGGCTGCTGCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACA 1315
Db 1262 GACAGTTAGTCATGGGGTGCTGCTGGGCTTTTAGAAGGCACAAGATAACATCTATTTATA 1321
Qy 1316 AGAAGCCTGGTACGCAAACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTC 1375
Db 1322 AGCGCCCAGATACCCAAACCATCATCAAAGTGAACAGCGATTTCCACTCATTCGTGCTGC 1381
Qy 1376 CACGCCTTACCAGCCACGGGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTG 1435
Db 1382 CCAGGATAGGCAGTAACACACTGGAGATCGGGCTGAGAACGAGAATCAGGAAAATGCTAG 1441
Qy 1436 AACCAACTGTCAAACCCGCACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCT 1495
Db 1442 AAGAGCACAAGGAGCCGTCACCTCTCATTACTGCCGAGGACATACAAGAGGCTAAGTGCG 1501
Qy 1496 TACAGCAAGAAGCTGAAGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCT 1555
Db 1502 CAGCCGATGAGGCTAAGGAAGTGCGTGAAGCCGAGGAGCTGCGCGCTGCACTACCACCTT 1561
Qy 1556 TGCTCCCTGAAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGG 1615
Db 1562 TGGCAGCTGATTTTGAGGAGCCCACTCTGGAAGCCGATGTCGACTTGATGTTACAAGAGG 1621
Qy 1616 CAGGAGCAGGTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCG 1675
Db 1622 CTGGGGCCGGCTCAGTGGAGACACCTCGTGGCTTGATAAAGGTTACCAGCTATGCCGGCG 1681
Qy 1676 AAGAGAAGATTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAAC 1735
Db 1682 AGGACAAGATCGGCTCTTACGCAGTGCTTTCTCCACAGGCTGTACTCAAGAGTGAGAAAC 1741
Qy 1736 TGGCGTGTATCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAG 1795
Db 1742 TATCTTGCATTCACCCTCTCGCTGAACAAGTCATAGTGATAACACACTCTGGCCGAAAAG 1801
Qy 1796 GGAGATACAAAGTCGAGCCATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCC 1855
Db 1802 GGCGTTATGCCGTGGAACCCTACCATGGAAAAGTAGTGGTGCCAGAGGGACATGCAATAC 1861
Qy 1856 CTGTTCAAGACTTCCAGGCATTGAGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGT 1915
Db 1862 CCGTCCAGGACTTTCAAGCTCTGAGTGAAAGTGCCACCATCGTGTACAACGAACGAGAGT 1921
Qy 1916 TCGTAAACAGATACCTGCACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAG 1975
Db 1922 TCGTAAACAGGTACCTGCACCATATTGCCACACATGGAGGAGCGCTGAACACAGATGAAG 1981
Qy 1976 AGTACTATAAGACTGTAAAGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGACG 2035
Db 1982 AATATTACAAAACTGTCAAGCCCAGCGAGCACGACGGCGAATACCTGTATGACATCGACA 2041
Qy 2036 CACGAAAGTGTGTTAAGCGAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAG 2095
Db 2042 GGAAACAATGCGTCAAGAAAGAATTAGTCACTGGGCTAGGGCTTACAGGCGAGCTGGTGG 2101
Qy 2096 ATCCACCATTTCACGAGTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACA 2155
Db 2102 ATCCTCCCTTCCATGAATTTGCCTACGAGAGTCTGAGAACACGTCCGGCCGCTCCTTACC 2161
Qy 2156 AAGTCCCAACCATCGGAGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAA 2215



```
Qy       4373 CATTGCATCACCTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCT 4432
Db       4382 CATTGAACCATTTGCTGACAGCTTTAGACACCACTGATGCAGATGTAGCCATATACTGCA 4441

Qy       4433 TGGATAAACAATGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAA 4492
Db       4442 GGGACAAGAAATGGGAAATGACTCTCAAGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGG 4501

Qy       4493 TTCTGGA---------------TGATGACAAGCCAGTAGACATTGACTTGGTCAGGGTCC 4537
Db       4502 AGATATGCATATCAGACGACTCTTCGGTGACAGAACCGGATGCAGAGCTGGTGAGGGTAC 4561

Qy       4538 ACCCAAACAGCTCTTTGGCAGGCAGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTATT 4597
Db       4562 ATCCGAAGAGTTCTTTGGCTGGAAGGAAGGGCTACAGCACAAGTGATGGCAAGACTTTCT 4621

Qy       4598 CATACCTGGAAGGTACACGATTCCATCAGACCGCCAAGGACATTGCCGAAATCCATGCAA 4657
Db       4622 CATATTTGGAAGGGACCAAATTTCACCAGGCGGCCAAGGATATAGCAGAAATTAATGCCA 4681

Qy       4658 TGTGGCCCAACAAATCTGAGGCTAATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTA 4717
Db       4682 TGTGGCCAGTTGCAACGGAGGCCAATGAGCAAGTATGCATGTATATCCTCGGTGAAAGCA 4741

Qy       4718 TGTCCAGCATCCGCTCCAAATGCCCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACA 4777
Db       4742 TGAGCAGCATTAGGTCGAAATGCCCCGTCGAGGAGTCGGAAGCCTCCACACCACCTAGCA 4801

Qy       4778 CACTTCCATGCCTGTGTAATTACGCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTG 4837
Db       4802 CGCTGCCTTGCTTGTGCATCCATGCTATGACTCCAGAAAGAGTACAACGCCTAAAAGCCT 4861

Qy       4838 CGAAGAAAGAACAGTTCGCCGTATGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAG 4897
Db       4862 CACGTCCAGAACAAATTACTGTGTGCTCATCCTTTCCATTGCCGAAGTATAGAATCACTG 4921

Qy       4898 GCGTGCAGAAGCTACAGTGCAGCAAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTG 4957
Db       4922 GTGTGCAGAAGATCCAGTGCTCCCAGCCTATACTGTTCTCACCGAAGGTGCCTGCGTACA 4981

Qy       4958 TACACCCCAGGAAGTAC 4974
Db       4982 TTCATCCACGGAAGTAC 4998
```

Search completed: November 15, 2003, 17:26:38
Job time : 17550 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 22:29:23 ; Search time 22769 Seconds
(without alignments)
11651.760 Million cell updates/sec

Title: US-10-023-649-1_COPY_5000_11484
Perfect score: 6485
Sequence: 1 caccgccagcaacgacaacc.....tattttgtttttaaaatttc 6485

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6485	100.0	11484	14	AF214040	AF214040 Western e
2	4075.6	62.8	4170	14	EEWSP	J03854 Western equ
3	3850.2	59.4	4047	14	AF229608	AF229608 Western e
4	2705.6	41.7	4692	14	AF339476	AF339476 Highlands
5	2397.6	37.0	11663	6	AR042411	AR042411 Sequence
6	2397.6	37.0	11663	6	AR096557	AR096557 Sequence
7	2397.6	37.0	11663	6	BD071122	BD071122 System fo
8	2396	36.9	11663	6	AR042409	AR042409 Sequence
9	2396	36.9	11663	6	AR096555	AR096555 Sequence
10	2396	36.9	11663	6	BD071120	BD071120 System fo
11	2396	36.9	11663	6	I46902	I46902 Sequence 1
12	2396	36.9	11663	14	ACU38305	U38305 Sindbis-lik
13	2388.4	36.8	11703	6	AR042412	AR042412 Sequence
14	2388.4	36.8	11703	6	AR096558	AR096558 Sequence
15	2388.4	36.8	11703	6	BD071123	BD071123 System fo
16	2386.4	36.8	11717	6	AR042410	AR042410 Sequence
17	2386.4	36.8	11717	6	AR096556	AR096556 Sequence
18	2386.4	36.8	11717	6	BD071121	BD071121 System fo
19	2386.4	36.8	11717	14	ACU38304	U38304 Sindbis-lik
20	2380.4	36.7	11703	14	SINCG	J02363 Sindbis vir
21	2380.4	36.7	11740	6	AR221250	AR221250 Sequence
22	2380.4	36.7	11740	6	AR230677	AR230677 Sequence
23	2380.4	36.7	11740	6	AR234151	AR234151 Sequence
24	2380.4	36.7	11740	6	AR237761	AR237761 Sequence
25	2380.4	36.7	11740	6	BD078845	BD078845 Recombine
26	2380.4	36.7	11740	6	BD085672	BD085672 Recombina
27	2372.2	36.6	11717	14	AF103734	AF103734 Sindbis-l
28	2369.2	36.5	16656	6	AR020969	AR020969 Sequence
29	2369.2	36.5	16656	6	AR043384	AR043384 Sequence
30	2369.2	36.5	16656	6	AR062299	AR062299 Sequence
31	2369.2	36.5	16656	6	AR183758	AR183758 Sequence
32	2365.6	36.5	11708	14	SINOCK82	M69205 Ockelbo vir
33	2340.2	36.1	11569	14	AF429428	AF429428 Sindbis v
34	2329.4	35.9	11626	14	AF103728	AF103728 Sindbis v
35	2322.4	35.8	16958	6	AX665480	AX665480 Sequence
36	2227.4	34.3	11678	14	U01034	U01034 Eastern equ
37	2225.2	34.3	11687	6	AX286761	AX286761 Sequence
38	2214.2	34.1	11675	14	EEEVIRNA	X63135 Eastern Equ
39	2198.4	33.9	11684	6	AX286760	AX286760 Sequence
40	2154	33.2	4541	14	AF339474	AF339474 Buggy Cre
41	2150.8	33.2	4525	14	AF339475	AF339475 Fort Morg
42	2122.6	32.7	13905	6	AR137241	AR137241 Sequence
43	2122.6	32.7	13905	6	AR282856	AR282856 Sequence
44	1880	29.0	11824	14	AF126284	AF126284 Aura viru
45	1825.2	28.1	4676	14	AF339477	AF339477 Babanki v

ALIGNMENTS

RESULT 1
AF214040
LOCUS AF214040 11484 bp RNA linear VRL 27-JAN-2000
DEFINITION Western equine encephalomyelitis virus strain 71V-1658 nonstructural polyprotein and structural polyprotein genes, complete cds.
ACCESSION AF214040 AF143811
VERSION AF214040.1 GI:6760410
KEYWORDS .
SOURCE Western equine encephalomyelitis virus
ORGANISM Western equine encephalomyelitis virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae; Alphavirus; WEEV complex.
REFERENCE 1 (bases 1 to 11484)
AUTHORS Netolitzky,D.J., Schmaltz,F.L., Parker,M.D., Rayner,G.A.,

```
          Fisher,G.R., Trent,D.W., Bader,D.E. and Nagata,L.P.
  TITLE   Complete genomic RNA sequence of western equine encephalitis virus
          and expression of the structural genes
  JOURNAL J. Gen. Virol. 81 (Pt 1), 151-159 (2000)
  MEDLINE 20109043
   PUBMED 10640553
REFERENCE 2  (bases 1 to 11484)
  AUTHORS Netolitzky,D.J., Schmaltz,F.L., Parker,M.D., Rayner,G.A.,
          Fisher,G.R., Trent,D.W., Bader,D.E. and Nagata,L.P.
  TITLE   Direct Submission
  JOURNAL Submitted (08-DEC-1999) Medical Countermeasures Section, Defence
          Research Establishment Suffield, P.O. Box 4000, Stn Main, Medicine
          Hat, Alberta T1A 8K6, Canada
FEATURES             Location/Qualifiers
     source          1. .11484
                     /organism="Western equine encephalomyelitis virus"
                     /mol_type="genomic RNA"
                     /strain="71V-1658"
                     /db_xref="taxon:11039"
     misc_feature    1. .11484
                     /note="sequence of this entry added to the sequence of
                     AF214039, strain CBA87, constitutes the Western equine
                     encephalomyelitis virus complete genome"
     5'UTR           <1. .24
     CDS             25. .7428
                     /note="encodes nsP1, nsP2, nsP3 and nsP4 proteins"
                     /codon_start=1
                     /transl_except=(pos:5581. .5583,aa:OTHER)
                     /product="nonstructural polyprotein"
                     /protein_id="AAF28339.1"
                     /db_xref="GI:6760411"
                     /translation="MERIHVDLDADSPYVKSLQRTFPQFEIEARQVTDNDHANARAFS
                     HVATKLIESEVDRDQVILDIGSAPVRHAHSNHRYHCICPMISAEDPDRLQRYAERLKK
                     SDITDKNIASKAADLLEVMSTPDAETPSLCMHTDATCRYFGSVAVYQDVYAVHAPTSI
                     YHQALKGVRTIYWIGFDTTPFMYKNMAGSYPTYNTNWADERVLEARNIGLGNSDLQES
                     RLGKLSILRKKRLQPTNKIIFSVGSTIYTEDRSLLRSWHLPNVFHLKGKSNFTGRCGT
                     IVSCEGYVIKKITISPGLYGKVENLASTMHREGFLSCKVTDTLRGERVSFAVCTYVPA
                     TLCDQMTGILATDVSVDDAQKLLVGLNQRIVVNGRTQRNTNTMQNYLLPVVAQAFSRW
                     AREHRADLDDEKELGVRERTLTMGCCWAFKTQKITSIYKKPGTQTIKKVPAVFDSFVI
                     PRLTSHGLDMGFRRRLKLLLEPTVKPAPAITMADVEHLRGLQQEAEEVAAAEEIREAL
                     PPLLPEIEKETVEAEVDLIMQEAGAGSVETPRGHIRVTSYPGEEKIGSYAILSPQAVL
                     NSEKLACIHPLAEQVLVMTHKGRAGRYKVEPYHGKVIVPEGTAVPVQDFQALSESATI
                     VFNEREFVNRYLHHIAINGGALNTDEEYYKTVKTQDTDSEYVFDIDARKCVKREDAGP
                     LCLTGDLVDPPFHEFAYESLKTRPAAPHKVPTIGVYGVPGSGKSGIIKSAVTKKDLVV
                     SAKKENCAEIIRDVRRMRRMDVAARTVDSVLLNGVKHPVNTLYIDEAFACHAGTLLAL
                     IAIVKPKKVVLCGDPKQCGFFNMMCLKVHFNHDICTEVYHKSISRRCTQTVTAIVSTL
                     FYDKRMKTVNPCADKIIIDTTGTTKPHKDDLILTCFRGWVKQLQIDYKNHEIMTAAAS
                     QGLTRKGVYAVRYKVNENPLYSQTSEHVNVLLTRTEKRIVWKTLAGDPWIKTLTAKYP
                     GDFTASLDDWQREHDAIMARVLDKPQTADVFQNKVNVCWAKALEPVLATANIVLTRQQ
                     WETLHPFKHDRAYSPEMALNFFCTRFFGVDLDSGLFSAPTVALTYRDQHWDNSPGKNM
                     YGLNREVAKELSRRYPCITKAVDTGRVADIRNNTIKDYSPTINVVPLNRRLPHSLIVD
                     HKGQGTTDHSGFLSKMKGKSVLVIGDPISIPGKKVESMGPLPTNTIRCDLDLGIPSHV
                     GKYDIIFVNVRTPYRNHHYQQCEDHAIHHSMLTCKAVHHLNTGGTCVAIGYGLADRAT
                     ENIITAVARSFRFTRVCQPKNTAENTEVLFVFFGKDNGNHTHDQDRLGVVLDNIYQGS
                     TRYEAGRAPAYRVIRGDISKSADQAIVNAANSKGQPGSGVCGALYRKWPAAFDRQPIA
                     VGTARLVKHEPLIIHAVGPNFSKMPEPEGDLKLAAAYMSIASIVNAERITKISVPLLS
                     TGIYSGGKDRVMQSLHHLFTAFDTTDADVTIYCLDKQWETRIIEAIHRKESVEILDDD
                     KPVDIDLVRVHPNSSLAGRPGYSVNEGKLYSYLEGTRFHQTAKDIAEIHAMWPNKSEA
                     NEQICLYILGESMSSIRSKCPVEESEASAPPHTLPCLCNYAMTAERVYRLRSAKKEQF
                     AVCSSFLLPKYRITGVQKLQCSKPVLFSGVVPPAVHPRKYAEIILETPPPPATTTVIC
                     EPTVPERIPSPVISRAPSAESLLSLGGVSFSSSATRSSTAWSDYDRRFVVTADVHQAN
                     TSTWSIPSAPGLDVQLPSDVTDSHWSIPSASGFEVRTPSVQDLTAECAKPRGLAEIMQ
                     DFNTAPFQFLSDYRPVPAPRRRPIPSPRSTASAPPVPKPRRTKYQQPPGVARAISEAE
                     LDEYIRQHSNXRYEAGAYIFSSETGQGHLQQKSVRQCKLQEPILDRAVHEKYYAPRLD
                     LEREKMLQKKLQLCASEGNRSRYQSRKVENMKAITAERLISGLGTYLSSEVNPVECYR
                     VNYPVPIYSSTVINRFTSAEVAVKTCNLVIQENYPTVASYCITDEYDAYLDMVDGASC
                     CLDTATFCPAKLRSYPKKHSYLQPEIRSAVPSPIQNTLQNVLAAAATKRNCNVTQMREL
                     PVLDSAAFNVDCFKKYACNDEYWDTFRDNPIRLTTENVTQYVTKLKGPKAAALFANTH
                     NLKPLQEIPMDQFVMDLKRDVKVTPGTKHTEERPKVQVIQAADPLATAYLCGIHRELV
                     RRLNAVLLPNIHTLFDMSAEDFDAIIAEHFHHGDPVLETDIASFDKSEDDAIAISALM
                     ILEDLGVDQPLLDLIEAAFGNITSVHLPTGTRFKFGAMMKSGMFLTLFVNTLVNIMIA
                     SRVLRERLTTSACAASIGDDNIVHGVVSDTLMAERCATWLNMEVKIIDAVIGIKAPYF
                     CGGFILVDQITGTACRVADPLKRLFKLGKPLPVDDTQDCDRRRALHDEAMRWNRIGIT
                     DELVKAVESRYEIILAGLIITSLSTLAESVKNFKSIRGSPITLYG"
     CDS             7473. .11183
                     /note="encodes nucleocapsid, E3, E2, 6K and E1 proteins"
                     /codon_start=1
                     /product="structural polyprotein"
                     /protein_id="AAF28340.1"
                     /db_xref="GI:6760412"
                     /translation="MFPYPQLNFPPVYPTNPMAYRDPNPPRRRWRPFRPPLAAQIEDL
                     RRSIVNLTFKQRSPNPPPGPPPKKKKSAPKPKPTQPKKKKQQAKRTKRKPKPGKRQRM
                     CMKLESDKTFPIMLNGQVNGYACVVGGRLMKPLHVEGKIDNEQLAAVKLKKASMYDLE
                     YGDVPQNMKSDTLQYTSDKPPGFYNWHHGAVQYENGRFTVPRGVGGKGDSGRPILDNR
                     GRVVAIVLGGANEGTRTALSVVTWNQKGVTIRDTPEGSEPWSLVTALCVLSNVTFPCD
                     KPPVCYSLTPERTLDVLEENVDNPNYDTLLENVLKCPSRRPKRSITDDFTLTSPYLGF
                     CPYCRHSTPCFSPIKIENVWDESDDGSIRIQVSAQFGYNQAGTADVTKFRYMSFDHDH
                     DIKEDSMEKIAISTSGPCRRLGHKGYFLLAQCPPGDSVTVSITSGASENSCTVEKKIR
                     RKFVGREEYLFPPVHGKLVKCHVYDHLKETSAGYITMHRPGPHAYKSYLEEASGEVYI
                     KPPSGKNVTYECKCGDYSTGIVSTRTKMNGCTKAKQCIAYKSDQTKWVFNSPDLIRHT
                     DHSVQGKLHIPFRLTPTVCPVPLAHTPTVTKWFKGITLHLTAMRPTLLTTRKLGLRAD
                     ATAEWITGSTSRNFSVGREGLEYVWGNHEPVRVWAQESAPGDPHGWPHEIIIHYYHRH
                     PVYTVIVLCGVALAILVGTASSAACIAKARRDCLTPYALAPNATVPTALAVLCCIRPT
                     NAETFGETLNHLWFNNQPFLWAQLCIPLAALVILFRCFSCCMPFLLVAGVCLGKVDAF
                     EHATTVPNVPGIPYKALVERAGYAPLNLEITVVSSELTPSTNKEYVTCKFHTVIPSPQ
                     VKCCGSLECKASSKADYTCRVFGGVYPFMWGGAQCFCDSENTQLSEAYVEFAPDCTID
                     HAVALKVHTAALKVGLRIVYGNTTAHLDTFVNGVTPGSSRDLKVIAGPISAAFSPFDH
                     KVVIRKGLVYNYDFPEYGAMKPGAFGDIQASSLDATDIVARTDIRLLKPSVKNIHVPY
                     TQAVSGYEMWKNNSGRPLQETAPFGCKIEVEPLRASNCAYGHIPISIDIPDAAFVRSS
                     ESPTILEVSCTVADCIYSADFGGSLTLQYKADREGHCPVHSHSTTAVLKEATTHVTAV
                     GSITLHFSTSSPQANFIVSLCGKKSTCNAECKPPADHIIGEPHKVDQEFQAAVSKTSW
                     NWLLALFGGASSLIVVGLIVLVCSSMLINTRR"
     3'UTR           11184. .11484
BASE COUNT     3251 a    2842 c    2808 g    2583 t
ORIGIN

  Query Match           100.0%;  Score 6485;  DB 14;  Length 11484;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 6485;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CACCGCCAGCAACGACAACCGTAATATGTGAACCAACTGTGCCAGAACGTATACCCAGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5000 CACCGCCAGCAACGACAACCGTAATATGTGAACCAACTGTGCCAGAACGTATACCCAGTC 5059

Qy         61 CGGTGATTTCTAGAGCACCAAGTGCGGAATCACTGCTATCGCTTGGCGGCGTCTCGTTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5060 CGGTGATTTCTAGAGCACCAAGTGCGGAATCACTGCTATCGCTTGGCGGCGTCTCGTTCT 5119

Qy        121 CTAGCTCTGCCACACGCTCGTCAACCGCCTGGAGCGACTATGACAGGCGGTTTGTGGTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5120 CTAGCTCTGCCACACGCTCGTCAACCGCCTGGAGCGACTATGACAGGCGGTTTGTGGTTA 5179

Qy        181 CAGCTGATGTGCATCAAGCGAACACGTCTACGTGGAGCATCCCTAGTGCTCCTGGCTTGG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5180 CAGCTGATGTGCATCAAGCGAACACGTCTACGTGGAGCATCCCTAGTGCTCCTGGCTTGG 5239

Qy        241 ACGTCCAGCTGCCTTCTGACGTCACTGATTCCCACTGGAGTATTCCAAGTGCATCAGGCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5240 ACGTCCAGCTGCCTTCTGACGTCACTGATTCCCACTGGAGTATTCCAAGTGCATCAGGCT 5299

Qy        301 TTGAAGTGAGAACACCATCTGTACAGGACCTAACTGCGGAGTGTGCGAAGCCTCGTGGAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5300 TTGAAGTGAGAACACCATCTGTACAGGACCTAACTGCGGAGTGTGCGAAGCCTCGTGGAC 5359

Qy        361 TGGCCGAAATAATGCAAGACTTCAATACTGCTCCTTTCCAGTTTCTTTCGGACTACAGAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5360 TGGCCGAAATAATGCAAGACTTCAATACTGCTCCTTTCCAGTTTCTTTCGGACTACAGAC 5419

Qy        421 CAGTACCGGCACCACGGAGACGCCCCATCCCATCACCTAGATCGACGGCTTCCGCACCTC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5420 CAGTACCGGCACCACGGAGACGCCCCATCCCATCACCTAGATCGACGGCTTCCGCACCTC 5479

Qy        481 CAGTTCCAAAGCCACGCAGGACTAAGTACCAACAACCACCAGGAGTCGCTAGAGCGATCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5480 CAGTTCCAAAGCCACGCAGGACTAAGTACCAACAACCACCAGGAGTCGCTAGAGCGATCT 5539
```

```
Qy       541 CAGAAGCGGAGTTGGACGAGTACATCCGTCAACACTCCAACTGACGGTATGAAGCGGGAG 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5540 CAGAAGCGGAGTTGGACGAGTACATCCGTCAACACTCCAACTGACGGTATGAAGCGGGAG 5599

Qy       601 CGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTCAACAGAAATCAGTACGTCAAT 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5600 CGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTCAACAGAAATCAGTACGTCAAT 5659

Qy       661 GTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATGAGAAGTATTACGCCCCGCGCC 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5660 GTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATGAGAAGTATTACGCCCCGCGCC 5719

Qy       721 TCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGCAATTATGCGCCTCTGAAGGAA 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5720 TCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGCAATTATGCGCCTCTGAAGGAA 5779

Qy       781 ATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGAAAGCAATTACAGCGGAGCGAC 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5780 ATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGAAAGCAATTACAGCGGAGCGAC 5839

Qy       841 TCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGAATCCTGTCGAGTGTTACAGAG 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5840 TCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGAATCCTGTCGAGTGTTACAGAG 5899

Qy       901 TCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTAACAGGTTTACATCTGCAGAGG 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5900 TCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTAACAGGTTTACATCTGCAGAGG 5959

Qy       961 TCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATTACCCTACAGTAGCCAGTTATT 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5960 TCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATTACCCTACAGTAGCCAGTTATT 6019

Qy      1021 GTATAACAGATGAATACGATGCGTATCTTGACATGGTGGACGGCGCATCGTGCTGTCTAG 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6020 GTATAACAGATGAATACGATGCGTATCTTGACATGGTGGACGGCGCATCGTGCTGTCTAG 6079

Qy      1081 ATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACCCAAAGAAGCATAGCTATTTGC 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6080 ATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACCCAAAGAAGCATAGCTATTTGC 6139

Qy      1141 AGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGAATACATTACAAAATGTATTGG 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6140 AGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGAATACATTACAAAATGTATTGG 6199

Qy      1201 CTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGCGAGAATTACCTGTCTTAGATT 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6200 CTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGCGAGAATTACCTGTCTTAGATT 6259

Qy      1261 CGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCATGCAATGATGAGTACTGGGATA 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6260 CGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCATGCAATGATGAGTACTGGGATA 6319

Qy      1321 CCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACGTTACGCAATATGTGACAAAGC 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6320 CCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACGTTACGCAATATGTGACAAAGC 6379

Qy      1381 TGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTCATAATCTAAAACCGTTGCAGG 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6380 TGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTCATAATCTAAAACCGTTGCAGG 6439

Qy      1441 AGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAGATGTCAAAGTTACTCCCGGCA 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6440 AGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAGATGTCAAAGTTACTCCCGGCA 6499

Qy      1501 CGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTCAGGCTGCAGATCCCCTTGCTA 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6500 CGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTCAGGCTGCAGATCCCCTTGCTA 6559

Qy      1561 CCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTAGACTGAATGCGGTGCTTCTGC 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6560 CCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTAGACTGAATGCGGTGCTTCTGC 6619

Qy      1621 CAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATTTTGATGCGATTATTGCTGAAC 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6620 CAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATTTTGATGCGATTATTGCTGAAC 6679

Qy      1681 ATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCGCGTCGTTTGATAAAAGCGAAG 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6680 ATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCGCGTCGTTTGATAAAAGCGAAG 6739

Qy      1741 ACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGGACTTAGGTGTCGACCAACCGC 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6740 ACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGGACTTAGGTGTCGACCAACCGC 6799

Qy      1801 TCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACATCTGTGCACCTACCTACAGGAA 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6800 TCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACATCTGTGCACCTACCTACAGGAA 6859

Qy      1861 CGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGTTCTTAACGCTGTTTGTCAACA 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6860 CGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGTTCTTAACGCTGTTTGTCAACA 6919

Qy      1921 CACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTGAACGGTTAACCACGTCAGCGT 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6920 CACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTGAACGGTTAACCACGTCAGCGT 6979

Qy      1981 GCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTGTCGTCTCCGACACCTTGATGG 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6980 GCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTGTCGTCTCCGACACCTTGATGG 7039

Qy      2041 CGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAATTATTGATGCAGTTATTGGTA 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7040 CGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAATTATTGATGCAGTTATTGGTA 7099

Qy      2101 TCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGGACCAGATAACAGGCACAGCCT 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7100 TCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGGACCAGATAACAGGCACAGCCT 7159

Qy      2161 GCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTGGAAAACCATTGCCAGTCGATG 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7160 GCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTGGAAAACCATTGCCAGTCGATG 7219

Qy      2221 ATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATGAAGCAATGCGATGGAACAGAA 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7220 ATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATGAAGCAATGCGATGGAACAGAA 7279

Qy      2281 TTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCAGATACGAGATCATACTGGCAG 2340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7280 TTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCAGATACGAGATCATACTGGCAG 7339

Qy      2341 GCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCGTTAAGAACTTCAAGAGCATAA 2400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7340 GCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCGTTAAGAACTTCAAGAGCATAA 7399

Qy      2401 GAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGGTGACGTAGTAGACACGCACCT 2460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7400 GAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGGTGACGTAGTAGACACGCACCT 7459

Qy      2461 ACCCACCGGCAGAATGTTTCCATACCCTCAGCTGAACTTTCCACCAGTTTACCCTACAAA 2520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7460 ACCCACCGGCAGAATGTTTCCATACCCTCAGCTGAACTTTCCACCAGTTTACCCTACAAA 7519

Qy      2521 TCCGATGGCTTACCGAGATCCAAACCCTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCC 2580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7520 TCCGATGGCTTACCGAGATCCAAACCCTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCC 7579

Qy      2581 GCTGGCTGCTCAAATCGAAGATCTTAGGAGGTCGATAGTCAACTTGACTTTCAAACAACG 2640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7580 GCTGGCTGCTCAAATCGAAGATCTTAGGAGGTCGATAGTCAACTTGACTTTCAAACAACG 7639

Qy      2641 ATCACCTAATCCGCCGCCAGGTCCACCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAA 2700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7640 ATCACCTAATCCGCCGCCAGGTCCACCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAA 7699

Qy      2701 ACCTACTCAGCCTAAAAAGAAGAAGCAGCAAGCCAAGAGGACGAAACGCAAGCCTAAACC 2760
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

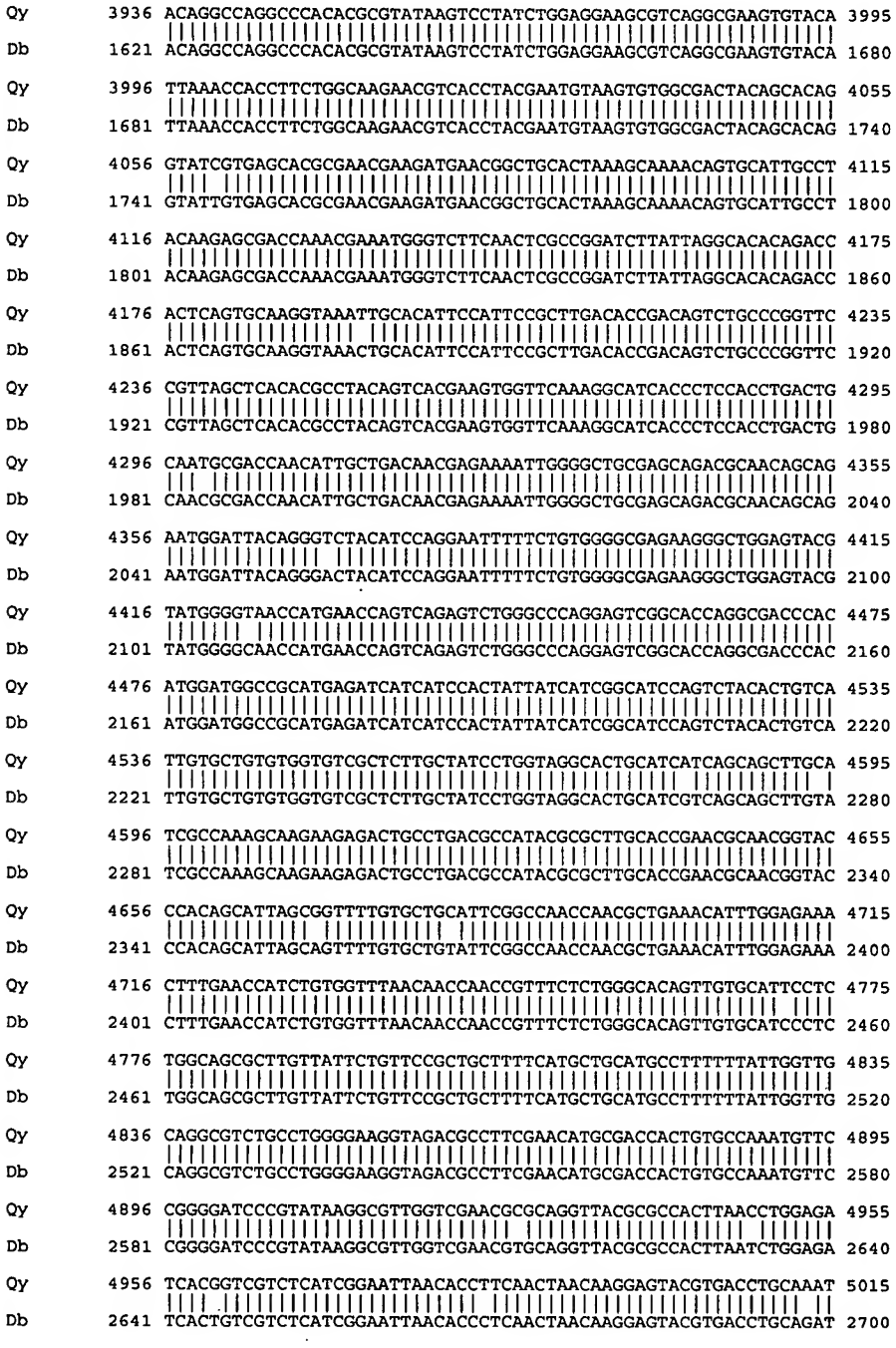
```
Db       7700 ACCTACTCAGCCTAAAAAGAAGAAGCAGCAAGCCAAGAGGACGAAACGCAAGCCTAAACC 7759
Qy       2761 AGGGAAACGACAACGTATGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATCATGCT 2820
Db       7760 AGGGAAACGACAACGTATGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATCATGCT 7819
Qy       2821 GAACGGCCAAGTGAATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACTCCA 2880
Db       7820 GAACGGCCAAGTGAATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACTCCA 7879
Qy       2881 CGTTGAAGGAAAAATTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAGCAT 2940
Db       7880 CGTTGAAGGAAAAATTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAGCAT 7939
Qy       2941 GTACGACTTGGAGTACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTACAC 3000
Db       7940 GTACGACTTGGAGTACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTACAC 7999
Qy       3001 CAGCGACAAACCACCGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAATGG 3060
Db       8000 CAGCGACAAACCACCGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAATGG 8059
Qy       3061 GAGATTTACCGTACCGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGA 3120
Db       8060 GAGATTTACCGTACCGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGA 8119
Qy       3121 CAACAGAGGCAGAGTTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGC 3180
Db       8120 CAACAGAGGCAGAGTTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGC 8179
Qy       3181 GCTTTCAGTGGTCACTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTC 3240
Db       8180 GCTTTCAGTGGTCACTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTC 8239
Qy       3241 TGAACCGTGGTCACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGA 3300
Db       8240 TGAACCGTGGTCACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGA 8299
Qy       3301 CAAACCACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAA 3360
Db       8300 CAAACCACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAA 8359
Qy       3361 CGTCGACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAATGTCCATCACGCCG 3420
Db       8360 CGTCGACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAATGTCCATCACGCCG 8419
Qy       3421 GCCCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCCC 3480
Db       8420 GCCCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCCC 8479
Qy       3481 GTATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACGA 3540
Db       8480 GTATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACGA 8539
Qy       3541 ATCTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAGG 3600
Db       8540 ATCTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAGG 8599
Qy       3601 CACTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAAGGA 3660
Db       8600 CACTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAAGGA 8659
Qy       3661 AGACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCACAA 3720
Db       8660 AGACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCACAA 8719
Qy       3721 AGGGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCACGAG 3780
Db       8720 AGGGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCACGAG 8779
Qy       3781 CGGAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTAG 3840
Db       8780 CGGAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTAG 8839

Qy       3841 AGAGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATCA 3900
Db       8840 AGAGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATCA 8899
Qy       3901 CTTGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAA 3960
Db       8900 CTTGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAA 8959
Qy       3961 GTCCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGT 4020
Db       8960 GTCCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGT 9019
Qy       4021 CACCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGAT 4080
Db       9020 CACCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGAT 9079
Qy       4081 GAACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGT 4140
Db       9080 GAACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGT 9139
Qy       4141 CTTCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACAT 4200
Db       9140 CTTCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACAT 9199
Qy       4201 TCCATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCAC 4260
Db       9200 TCCATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCAC 9259
Qy       4261 GAAGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAAC 4320
Db       9260 GAAGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAAC 9319
Qy       4321 GAGAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAG 4380
Db       9320 GAGAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAG 9379
Qy       4381 GAATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAG 4440
Db       9380 GAATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAG 9439
Qy       4441 AGTCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCAT 4500
Db       9440 AGTCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCAT 9499
Qy       4501 CCACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGC 4560
Db       9500 CCACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGC 9559
Qy       4561 TATCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCT 4620
Db       9560 TATCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCT 9619
Qy       4621 GACGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTG 4680
Db       9620 GACGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTG 9679
Qy       4681 CATTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAA 4740
Db       9680 CATTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAA 9739
Qy       4741 CCAACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCG 4800
Db       9740 CCAACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCG 9799
Qy       4801 CTGCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGA 4860
Db       9800 CTGCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGA 9859
Qy       4861 CGCCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGT 4920
Db       9860 CGCCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGT 9919
```

```
Qy       4921 CGAACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAAC 4980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9920 CGAACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAAC 9979

Qy       4981 ACCTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACA 5040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9980 ACCTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACA 10039

Qy       5041 AGTTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCG 5100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10040 AGTTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCG 10099

Qy       5101 CGTTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGA 5160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10100 CGTTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGA 10159

Qy       5161 GAACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGC 5220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10160 GAACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGC 10219

Qy       5221 AGTCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAA 5280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10220 AGTCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAA 10279

Qy       5281 CACCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCT 5340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10280 CACCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCT 10339

Qy       5341 GAAGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCAT 5400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10340 GAAGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCAT 10399

Qy       5401 CAGAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGC 5460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10400 CAGAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGC 10459

Qy       5461 GTTCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACAT 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10460 GTTCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACAT 10519

Qy       5521 ACGGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGG 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10520 ACGGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGG 10579

Qy       5581 GTATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATG 5640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10580 GTATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATG 10639

Qy       5641 TAAAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTC 5700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10640 TAAAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTC 10699

Qy       5701 GATTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGT 5760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10700 GATTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGT 10759

Qy       5761 TAGCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACA 5820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10760 TAGCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACA 10819

Qy       5821 GTACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTT 5880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10820 GTACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTT 10879

Qy       5881 GAAGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATC 5940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10880 GAAGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATC 10939

Qy       5941 GAGCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGA 6000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10940 GAGCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGA 10999

Qy       6001 ATGTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCA 6060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       11000 ATGTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCA 11059

Qy       6061 GGCGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATC 6120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       11060 GGCGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATC 11119

Qy       6121 CCTCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAG 6180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       11120 CCTCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAG 11179

Qy       6181 ATGACTGAGCGCGGACACTGACATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGG 6240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       11180 ATGACTGAGCGCGGACACTGACATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGG 11239

Qy       6241 TGCATAATGCCACGCGCCGCTTGACACTAAAACTCGATGTATTTCCGAGGAAGCACAGTG 6300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       11240 TGCATAATGCCACGCGCCGCTTGACACTAAAACTCGATGTATTTCCGAGGAAGCACAGTG 11299

Qy       6301 CATAATGCTGTGCAGTGTCACATTAATCGTATATCACACTACATATTAACAACACTATAT 6360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       11300 CATAATGCTGTGCAGTGTCACATTAATCGTATATCACACTACATATTAACAACACTATAT 11359

Qy       6361 CACTTTTATGAGACTCACTATGGGTCTCTAATATACACTACACATATTTTACTTAAAAAC 6420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       11360 CACTTTTATGAGACTCACTATGGGTCTCTAATATACACTACACATATTTTACTTAAAAAC 11419

Qy       6421 ACTATACACACTTTATAAATTCTTTTATAATTTTTCTTTTGTTTTTATTTTGTTTTTAAA 6480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       11420 ACTATACACACTTTATAAATTCTTTTATAATTTTTCTTTTGTTTTTATTTTGTTTTTAAA 11479

Qy       6481 ATTTC 6485
             |||||
Db       11480 ATTTC 11484
```

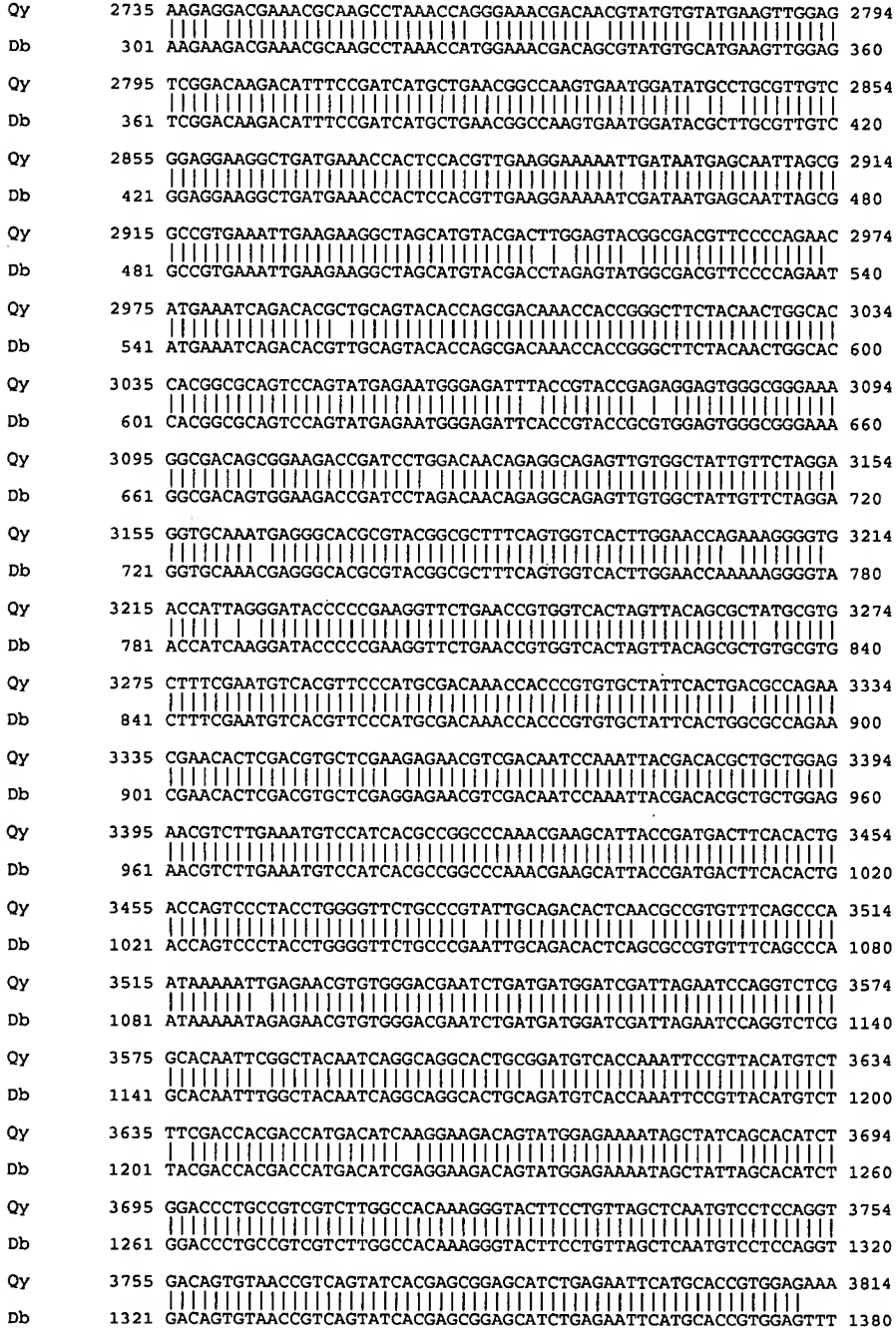
RESULT 2
EEWSP

```
LOCUS       EEWSP                   4170 bp ss-RNA     linear   VRL 02-AUG-1993
DEFINITION  Western equine encephalitis virus RNA encoding structural proteins
            at the 3' end of the genome.
ACCESSION   J03854
VERSION     J03854.1  GI:323728
KEYWORDS    6 kd protein; capsid protein; envelope protein.
SOURCE      Western equine encephalomyelitis virus
  ORGANISM  Western equine encephalomyelitis virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
            Alphavirus; WEEV complex.
REFERENCE   1  (bases 1 to 4170)
  AUTHORS   Hahn,C.S., Lustig,S., Strauss,E.G. and Strauss,J.H.
  TITLE     Western equine encephalitis virus is a recombinant virus
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 85 (16), 5997-6001 (1988)
  MEDLINE   88320369
   PUBMED   3413072
COMMENT     Original source text: Western equine encephalitis virus (strain
            BFS1703) RNA, passed in suckling mice and VERO cells, originally
            from Cx.tarsalis.
FEATURES             Location/Qualifiers
     source          1. .4170
                     /organism="Western equine encephalomyelitis virus"
                     /mol_type="genomic RNA"
                     /db_xref="taxon:11039"
     CDS             <1. .114
                     /note="nonstructural protein 4"
                     /codon_start=1
                     /protein_id="AAA42998.1"
                     /db_xref="GI:323729"
                     /translation="SRYEIILAGLIITSLSTLAESVKNFKSIRGNPITLYG"
     CDS             159. .3869
                     /note="envelope polyprotein"
                     /codon_start=1
                     /protein_id="AAA42999.1"
                     /db_xref="GI:323730"
```

```
                     /translation="MFPYPQLNFPPVYPTNPMAYRDPNPPRCRWRPFRPPLAAQIEDL
                     RRSIANLTFKQRSPNPPPGPPPKKKKSAPKPKPTQPKKKKQQAKKTKRKPKPGKRQRM
                     CMKLESDKTFPIMLNGQVNGYACVVGGRLMKPLHVEGKIDNEQLAAVKLKKASMYDLE
                     YGDVPQNMKSDTLQYTSDKPPGFYNWHHGAVQYENGRFTVPRGVGGKGDSGRPILDNR
                     GRVVAIVLGGANEGTRTALSVVTWNQKGVTIKDTPEGSEPWSLVTALCVLSNVTFPCD
                     KPPVCYSLAPERTLDVLEENVDNPNYDTLLENVLKCPSRRPKRSITDDFTLTSPYLGF
                     CPYCRHSAPCFSPIKIENVWDESDDGSIRIQVSAQFGYNQAGTADVTKFRYMSFDHDH
                     DIKEDSMDKIAISTSGPCRRLGHKGYFLLAQCPPGDSVTVSITSGASENSCTVEKKIR
                     RKFVGREEYLFPPVHGKLVKCHVYDHLKETSAGYITMHRPGPHAYKSYLEEASGEVYI
                     KPPSGKNVTYECKCGDYSTGIVSTRTKMNGCTKAKQCIAYKSDQTKWVFNSPDLIRHT
                     DHSVQGKLHIPFRLTPTVCPVPLAHTPTVTKWFKGITLHLTATRPTLLTTRKLGLRAD
                     ATAEWITGTTSRNFSVGREGLEYVWGNHEPVRVWAQESAPGDPHGWPHEIIIHYYHRH
                     PVYTVIVLCGVALAILVGTASSAACIAKARRDCLTPYALAPNATVPTALAVLCCIRPT
                     NAETFGETLNHLWFNNQPFLWAQLCIPLAALVILFRCFSCCMPFLLVAGVCLGKVDAF
                     EHATTVPNVPGIPYKALVERAGYAPLNLEITVVSSELTPSTNKEYVTCRFHTVIPSPQ
                     VKCCGSLECKASSKADYTCRVFGGVYPFMWGGAQCFCDSENTQLSEAYVEFAPDCTID
                     HAVALKVHTAALKVGLRIVYGNTTAHLDTFVNGVTPGSSRDLKVIAGPISAAFSPFDH
                     KVVIRKGLVYNYDFPEYGAMKPGAFGDIQASSLDATDIVARTDIRLLKPSVKNIHVPY
                     TQAVSGYEMWKNNSGRPLQETAPFGCKIEVEPLRASNCAYGHIPISIDIPDAAFVRSS
                     ESPTILEVSCTVADCIYSADFGGSLTLQYKADREGHCPVHSHSTTAVLKEATTHVTAV
                     GSITLHFSTSSPQANFIVSLCGKKTTCNAECKPPADHIIGEPHKVDQEFQAAVSKTSW
                     NWLLALFGGASSLIVVGLIVLVCSSMLINTRR"
     mat_peptide     159. .935
                     /product="capsid protein"
     mat_peptide     936. .1115
                     /product="envelope protein E3"
     mat_peptide     1116. .2384
                     /product="envelope glycoprotein E2"
     mat_peptide     2385. .2549
                     /product="6 kd protein"
     mat_peptide     2550. .3866
                     /product="envelope glycoprotein E1"
BASE COUNT     1157 a   1080 c    992 g    941 t
ORIGIN

  Query Match             62.8%;  Score 4075.6;  DB 14;  Length 4170;
  Best Local Similarity   98.6%;  Pred. No. 0;
  Matches 4111;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy       2316 TCCAGATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAA 2375
             ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db          1 TCCAGATACGAGATCATACTGGCAGGCCTGATCATCACGTCCCTGTCCACGTTAGCCGAA 60

Qy       2376 AGCGTTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAA 2435
             ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db         61 AGCGTTAAGAACTTCAAGAGCATAAGAGGGAACCCAATCACCCTCTACGGCTGACCTAAA 120

Qy       2436 TAGGTGACGTAGTAGACACGCACCTACCCACCGGCAGAATGTTTCCATACCCTCAGCTGA 2495
             |||||||||||||||||||||||||||||||| || ||||||||||||||||||||||||||
Db        121 TAGGTGACGTAGTAGACACGCACCTACCCACCGCCAAAATGTTTCCATACCCTCAGCTGA 180

Qy       2496 ACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACCCTCCTAGGC 2555
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACCCTCCTAGGT 240

Qy       2556 GCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTAGGAGGTCGA 2615
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTAGGAGGTCGA 300

Qy       2616 TAGTCAACTTGACTTTCAAACAACGATCACCTAATCCGCCGCCAGGTCCACCGCCAAAGA 2675
             ||| |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TAGCCAACTTAACTTTCAAACAACGATCACCTAATCCGCCGCCAGGTCCACCGCCAAAGA 360

Qy       2676 AGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGAAGAAGCAGCAAGCCA 2735
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGAAGAAGCAGCAAGCCA 420

Qy       2736 AGAGGACGAAACGCAAGCCTAAACCAGGGAAACGACAACGTATGTGTATGAAGTTGGAGT 2795
             ||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        421 AGAAGACGAAACGCAAGCCTAAACCAGGGAAACGACAGCGTATGTGTATGAAGTTGGAGT 480

Qy       2796 CGGACAAGACATTTCCGATCATGCTGAACGGCCAAGTGAATGGATATGCCTGCGTTGTCG 2855
             |||||||||||||||||||||||||||||||||||||||||||||| || ||||||||||

Db        481 CGGACAAGACATTTCCGATCATGCTGAACGGCCAAGTGAATGGATACGCTTGCGTTGTCG 540

Qy       2856 GAGGAAGGCTGATGAAACCACTCCACGTTGAAGGAAAAATTGATAATGAGCAATTAGCGG 2915
             |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db        541 GAGGAAGGCTGATGAAACCACTCCACGTTGAAGGAAAAATCGATAATGAGCAATTAGCGG 600

Qy       2916 CCGTGAAATTGAAGAAGGCTAGCATGTACGACTTGGAGTACGGCGACGTTCCCCAGAACA 2975
             |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db        601 CCGTGAAATTGAAGAAGGCTAGCATGTACGACTTGGAGTATGGCGACGTTCCCCAGAATA 660

Qy       2976 TGAAATCAGACACGCTGCAGTACACCAGCGACAAACCACCGGGCTTCTACAACTGGCACC 3035
             |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db        661 TGAAATCAGACACGCTGCAGTACACCAGCGACAAACCACCGGGCTTTTACAACTGGCACC 720

Qy       3036 ACGGCGCAGTCCAGTATGAGAATGGGAGATTTACCGTACCGAGAGGAGTGGGCGGGAAAG 3095
             |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db        721 ACGGCGCAGTCCAGTATGAGAATGGGAGATTCACCGTACCGAGAGGAGTGGGCGGGAAAG 780

Qy       3096 GCGACAGCGGAAGACCGATCCTGGACAACAGAGGCAGAGTTGTGGCTATTGTTCTAGGAG 3155
             ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCGACAGTGGAAGACCGATCCTGGACAACAGAGGCAGAGTTGTGGCTATTGTTCTAGGAG 840

Qy       3156 GTGCAAATGAGGGCACGCGTACGGCGCTTTCAGTGGTCACTTGGAACCAGAAAGGGGTGA 3215
             ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GTGCAAACGAGGGCACGCGTACGGCGCTTTCAGTGGTCACTTGGAACCAGAAAGGGGTGA 900

Qy       3216 CCATTAGGGATACCCCCGAAGGTTCTGAACCGTGGTCACTAGTTACAGCGCTATGCGTGC 3275
             |||| | |||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        901 CCATCAAGGATACCCCCGAAGGTTCTGAACCGTGGTCACTAGTTACAGCGCTGTGCGTGC 960

Qy       3276 TTTCGAATGTCACGTTCCCATGCGACAAACCACCCGTGTGCTATTCACTGACGCCAGAAC 3335
             |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        961 TTTCGAATGTCACGTTCCCATGCGACAAACCACCCGTGTGCTATTCACTGGCGCCAGAAC 1020

Qy       3336 GAACACTCGACGTGCTCGAAGAGAACGTCGACAATCCAAATTACGACACGCTGCTGGAGA 3395
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GAACACTCGACGTGCTCGAAGAGAACGTCGACAATCCAAATTACGACACGCTGCTGGAGA 1080

Qy       3396 ACGTCTTGAAATGTCCATCACGCCGGCCCAAACGAAGCATTACCGATGACTTCACACTGA 3455
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACGTCTTGAAATGTCCATCACGCCGGCCCAAACGAAGCATTACCGATGACTTCACACTGA 1140

Qy       3456 CCAGTCCCTACCTGGGGTTCTGCCCGTATTGCAGACACTCAACGCCGTGTTTCAGCCCAA 3515
             |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db       1141 CCAGTCCCTACCTGGGGTTCTGCCCGTATTGCAGACACTCAGCGCCGTGTTTCAGCCCAA 1200

Qy       3516 TAAAAATTGAGAACGTGTGGGACGAATCTGATGATGGATCGATTAGAATCCAGGTCTCGG 3575
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TAAAAATTGAGAACGTGTGGGACGAATCTGATGATGGATCGATTAGAATCCAGGTCTCGG 1260

Qy       3576 CACAATTCGGCTACAATCAGGCAGGCACTGCGGATGTCACCAAATTCCGTTACATGTCTT 3635
             |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       1261 CACAATTCGGCTACAATCAGGCAGGCACTGCAGATGTCACCAAATTCCGTTACATGTCTT 1320

Qy       3636 TCGACCACGACCATGACATCAAGGAAGACAGTATGGAGAAAATAGCTATCAGCACATCTG 3695
             |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db       1321 TCGACCACGACCATGACATCAAGGAAGACAGTATGGATAAAATAGCTATTAGCACATCTG 1380

Qy       3696 GACCCTGCCGTCGTCTTGGCCACAAAGGGTACTTCCTGTTAGCTCAATGTCCTCCAGGTG 3755
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GACCCTGCCGTCGTCTTGGCCACAAAGGGTACTTCCTGTTAGCTCAATGTCCTCCAGGTG 1440

Qy       3756 ACAGTGTAACCGTCAGTATCACGAGCGGAGCATCTGAGAATTCATGCACCGTGGAGAAAA 3815
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ACAGTGTAACCGTCAGTATCACGAGCGGAGCATCTGAGAATTCATGCACCGTGGAGAAAA 1500

Qy       3816 AGATCAGGAGGAAGTTTGTCGGTAGAGAGGAGTACTTGTTCCCACCCGTCCATGGAAAGC 3875
             |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db       1501 AGATCAGGAGGAAGTTTGTCGGTAGAGAGGAGTACTTGTTCCCACCTGTCCATGGAAAGC 1560

Qy       3876 TGGTAAAGTGCCACGTTTACGATCACTTGAAGGAGACGTCTGCCGGGTACATAACCATGC 3935
             |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db       1561 TGGTAAAGTGCCACGTTTACGATCACTTGAAGGAGACGTCTGCCGGATACATAACCATGC 1620
```


Qy 3936 ACAGGCCAGGCCCACACGCGTATAAGTCCTATCTGGAGGAAGCGTCAGGCGAAGTGTACA 3995
Db 1621 ACAGGCCAGGCCCACACGCGTATAAGTCCTATCTGGAGGAAGCGTCAGGCGAAGTGTACA 1680
Qy 3996 TTAAACCACCTTCTGGCAAGAACGTCACCTACGAATGTAAGTGTGGCGACTACAGCACAG 4055
Db 1681 TTAAACCACCTTCTGGCAAGAACGTCACCTACGAATGTAAGTGTGGCGACTACAGCACAG 1740
Qy 4056 GTATCGTGAGCACGCGAACGAAGATGAACGGCTGCACTAAAGCAAAACAGTGCATTGCCT 4115
Db 1741 GTATTGTGAGCACGCGAACGAAGATGAACGGCTGCACTAAAGCAAAACAGTGCATTGCCT 1800
Qy 4116 ACAAGAGCGACCAAACGAAATGGGTCTTCAACTCGCCGGATCTTATTAGGCACACAGACC 4175
Db 1801 ACAAGAGCGACCAAACGAAATGGGTCTTCAACTCGCCGGATCTTATTAGGCACACAGACC 1860
Qy 4176 ACTCAGTGCAAGGTAAATTGCACATTCCATTCCGCTTGACACCGACAGTCTGCCCGGTTC 4235
Db 1861 ACTCAGTGCAAGGTAAACTGCACATTCCATTCCGCTTGACACCGACAGTCTGCCCGGTTC 1920
Qy 4236 CGTTAGCTCACACGCCTACAGTCACGAAGTGGTTCAAAGGCATCACCCTCCACCTGACTG 4295
Db 1921 CGTTAGCTCACACGCCTACAGTCACGAAGTGGTTCAAAGGCATCACCCTCCACCTGACTG 1980
Qy 4296 CAATGCGACCAACATTGCTGACAACGAGAAAATTGGGGCTGCGAGCAGACGCAACAGCAG 4355
Db 1981 CAACGCGACCAACATTGCTGACAACGAGAAAATTGGGGCTGCGAGCAGACGCAACAGCAG 2040
Qy 4356 AATGGATTACAGGGTCTACATCCAGGAATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACG 4415
Db 2041 AATGGATTACAGGGACTACATCCAGGAATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACG 2100
Qy 4416 TATGGGGTAACCATGAACCAGTCAGAGTCTGGGCCCAGGAGTCGGCACCAGGCGACCCAC 4475
Db 2101 TATGGGGCAACCATGAACCAGTCAGAGTCTGGGCCCAGGAGTCGGCACCAGGCGACCCAC 2160
Qy 4476 ATGGATGGCCGCATGAGATCATCATCCACTATTATCATCGGCATCCAGTCTACACTGTCA 4535
Db 2161 ATGGATGGCCGCATGAGATCATCATCCACTATTATCATCGGCATCCAGTCTACACTGTCA 2220
Qy 4536 TTGTGCTGTGTGGTGTCGCTCTTGCTATCCTGGTAGGCACTGCATCATCAGCAGCTTGCA 4595
Db 2221 TTGTGCTGTGTGGTGTCGCTCTTGCTATCCTGGTAGGCACTGCATCGTCAGCAGCTTGTA 2280
Qy 4596 TCGCCAAAGCAAGAAGAGACTGCCTGACGCCATACGCGCTTGCACCGAACGCAACGGTAC 4655
Db 2281 TCGCCAAAGCAAGAAGAGACTGCCTGACGCCATACGCGCTTGCACCGAACGCAACGGTAC 2340
Qy 4656 CCACAGCATTAGCGGTTTTGTGCTGCATTCGGCCAACCAACGCTGAAACATTTGGAGAAA 4715
Db 2341 CCACAGCATTAGCAGTTTTGTGCTGTATTCGGCCAACCAACGCTGAAACATTTGGAGAAA 2400
Qy 4716 CTTTGAACCATCTGTGGTTTAACAACCAACCGTTTCTCTGGGCACAGTTGTGCATTCCTC 4775
Db 2401 CTTTGAACCATCTGTGGTTTAACAACCAACCGTTTCTCTGGGCACAGTTGTGCATCCCTC 2460
Qy 4776 TGGCAGCGCTTGTTATTCTGTTCCGCTGCTTTTCATGCTGCATGCCTTTTTTATTGGTTG 4835
Db 2461 TGGCAGCGCTTGTTATTCTGTTCCGCTGCTTTTCATGCTGCATGCCTTTTTTATTGGTTG 2520
Qy 4836 CAGGCGTCTGCCTGGGGAAGGTAGACGCCTTCGAACATGCGACCACTGTGCCAAATGTTC 4895
Db 2521 CAGGCGTCTGCCTGGGGAAGGTAGACGCCTTCGAACATGCGACCACTGTGCCAAATGTTC 2580
Qy 4896 CGGGGATCCCGTATAAGGCGTTGGTCGAACGCGCAGGTTACGCGCCACTTAACCTGGAGA 4955
Db 2581 CGGGGATCCCGTATAAGGCGTTGGTCGAACGTGCAGGTTACGCGCCACTTAATCTGGAGA 2640
Qy 4956 TCACGGTCGTCTCATCGGAATTAACACCTTCAACTAACAAGGAGTACGTGACCTGCAAAT 5015
Db 2641 TCACTGTCGTCTCATCGGAATTAACACCCTCAACTAACAAGGAGTACGTGACCTGCAGAT 2700
Qy 5016 TCCACACAGTCATTCCTTCACCACAAGTTAAATGCTGCGGGTCCCTCGAGTGCAAGGCAT 5075
Db 2701 TCCACACAGTCATTCCTTCACCACAAGTTAAATGCTGCGGGTCCCTCGAGTGTAAGGCAT 2760
Qy 5076 CCTCAAAGGCGGATTACACATGCCGCGTTTTTGGCGGTGTGTACCCTTTCATGTGGGGAG 5135
Db 2761 CCTCAAAAGCGGATTACACATGCCGCGTTTTTGGCGGTGTGTACCCTTTCATGTGGGGAG 2820
Qy 5136 GCGCACAATGCTTCTGTGACAGTGAGAACACACAACTGAGTGAGGCGTACGTCGAGTTCG 5195
Db 2821 GCGCACAGTGCTTCTGTGACAGTGAGAACACACAACTGAGTGAGGCATACGTCGAGTTCG 2880
Qy 5196 CTCCAGACTGCACTATAGATCACGCAGTCGCACTAAAAGTTCACACAGCTGCTCTGAAAG 5255
Db 2881 CTCCAGACTGCACTATTGATCATGCAGTCGCACTAAAAGTTCACACAGCTGCTCTGAAAG 2940
Qy 5256 TCGGCCTGCGTATAGTATACGGCAACACCACCGCGCACCTGGATACGTTTGTCAATGGCG 5315
Db 2941 TCGGCCTGCGTATAGTATACGGCAATACCACCGCGCACCTGGATACGTTCGTCAATGGCG 3000
Qy 5316 TCACGCCAGGTTCCTCACGGGACCTGAAGGTCATAGCAGGGCCGATATCAGCCGCTTTTT 5375
Db 3001 TCACGCCAGGTTCCTCACGGGACCTGAAGGTCATAGCAGGGCCGATATCAGCCGCTTTTT 3060
Qy 5376 CACCCTTTGACCATAAGGTCGTCATCAGAAAGGGGCTTGTTTACAACTACGACTTCCCTG 5435
Db 3061 CACCCTTTGACCATAAGGTCGTCATCAGAAAGGGGCTTGTTTACAACTACGACTTCCCTG 3120
Qy 5436 AGTATGGAGCTATGAAACCAGGAGCGTTCGGCGATATTCAAGCATCCTCGCTTGATGCTA 5495
Db 3121 AATATGGAGCTATGAAACCAGGAGCGTTCGGCGATATTCAAGCATCCTCGCTTGATGCCA 3180
Qy 5496 CAGACATAGTAGCCCGCACTGACATACGGCTGCTGAAGCCTTCTGTCAAGAACATCCACG 5555
Db 3181 CAGACATAGTAGCCCGCACTGACATACGGCTGCTGAAGCCTTCTGTCAAGAACATCCACG 3240
Qy 5556 TCCCCTACACCCAAGCAGTATCAGGGTATGAAATGTGGAAGAACAACTCAGGACGACCCC 5615
Db 3241 TCCCCTACACCCAAGCAGTATCAGGGTATGAAATGTGGAAGAACAACTCAGGACGACCCC 3300
Qy 5616 TGCAAGAAACAGCACCATTTGGATGTAAAATTGAAGTGGAGCCTCTGCGAGCGTCTAACT 5675
Db 3301 TGCAAGAAACAGCACCATTTGGATGTAAAATTGAAGTGGAGCCTCTGCGAGCGTCTAACT 3360
Qy 5676 GTGCTTACGGGCACATCCCTATCTCGATTGACATCCCTGATGCAGCTTTTGTGAGATCAT 5735
Db 3361 GTGCTTACGGGCACATCCCTATCTCGATTGACATCCCTGACGCAGCTTTTGTGAGATCAT 3420
Qy 5736 CAGAATCACCAACAATTTTAGAAGTTAGCTGCACAGTAGCAGACTGCATTTATTCTGCAG 5795
Db 3421 CAGAATCACCAACAATTTTAGAAGTTAGCTGCACAGTAGCAGACTGCATTTATTCTGCAG 3480
Qy 5796 ACTTTGGTGGTTCTCTAACATTACAGTACAAAGCTGACAGGGAGGGACATTGTCCAGTTC 5855
Db 3481 ACTTTGGTGGTTCGCTAACATTACAGTACAAAGCTGACAGGGAGGGACATTGTCCAGTTC 3540
Qy 5856 ACTCCCACTCCACGACAGCTGTTTTGAAGGAAGCGACCACACATGTGACTGCCGTAGGCA 5915
Db 3541 ACTCCCACTCCACAACAGCTGTTTTGAAGGAAGCGACCACACATGTGACTGCCGTAGGCA 3600
Qy 5916 GCATAACACTACATTTTAGCACATCGAGCCCACAAGCAAATTTTATAGTTTCGCTATGCG 5975
Db 3601 GCATAACACTACATTTTAGCACATCGAGCCCACAAGCAAATTTTATAGTTTCGCTATGCG 3660
Qy 5976 GCAAGAAGTCCACCTGCAATGCTGAATGTAAACCACCGGCCGACCACATAATTGGAGAAC 6035
Db 3661 GCAAGAAGACCACCTGCAATGCTGAATGTAAACCACCGGCCGACCACATAATTGGAGAAC 3720
Qy 6036 CACATAAAGTCGACCAAGAATTCCAGGCGGCAGTTTCCAAAACATCTTGGAACTGGCTGC 6095
Db 3721 CACATAAAGTCGACCAAGAATTCCAGGCGGCAGTTTCCAAAACATCTTGGAATTGGCTGC 3780
Qy 6096 TTGCACTGTTTGGGGGAGCATCATCCCTCATTGTTGTAGGACTTATAGTGTTGGTCTGCA 6155

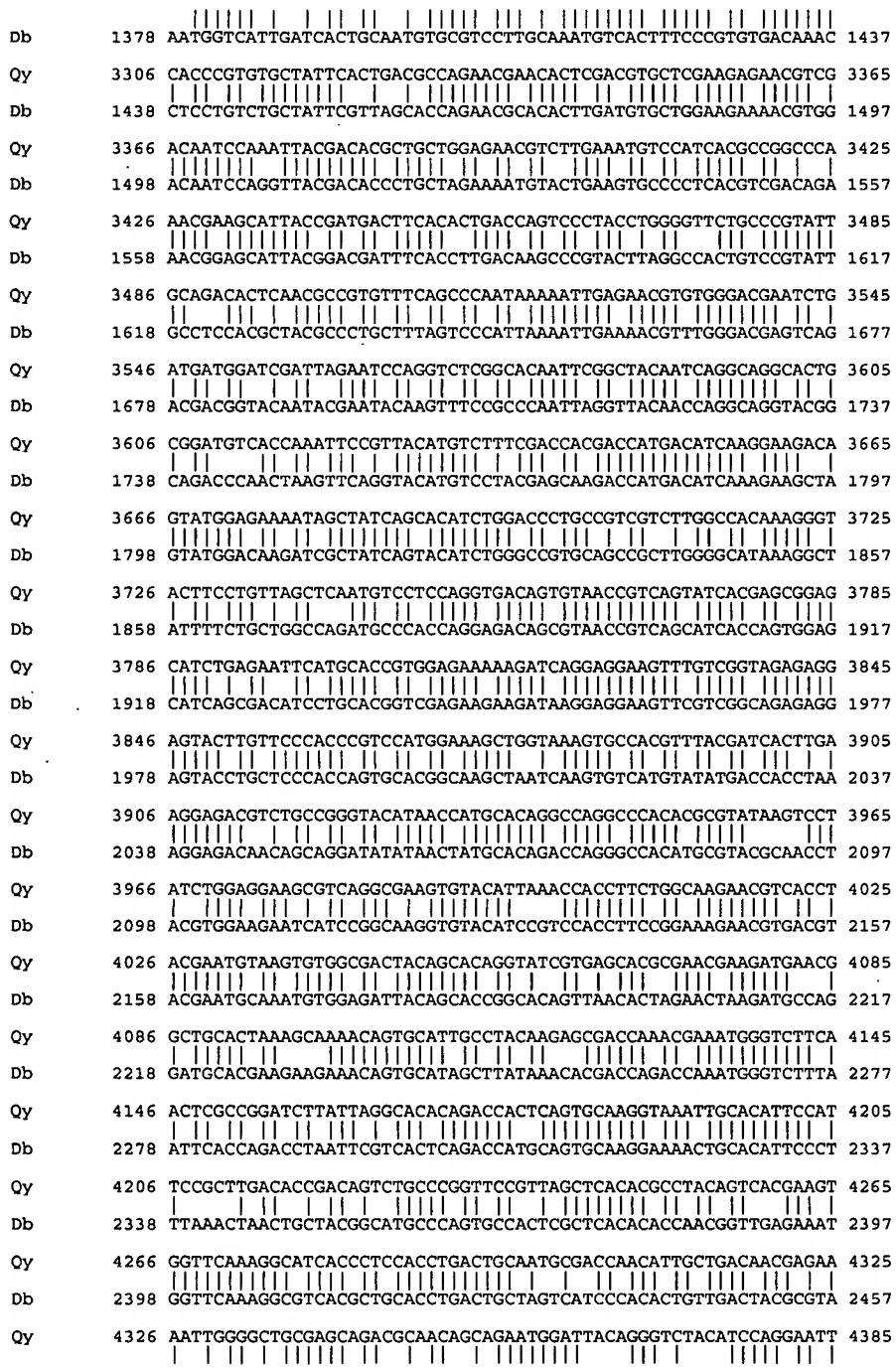
```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 TTGCACTGTTTGGGGGAGCATCATCCCTCATTGTTGTAGGACTTATAGTGTTGGTCTGCA 3840

Qy       6156 GCTCTATGCTTATAAACACACGTAGATGACTGAGCGCGGACACTGACATAGCGGTAAAAC 6215
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3841 GCTCTATGCTTATAAACACACGTAGATGACTGAGCGCGGACACTGACATAGCGGTAAAAC 3900

Qy       6216 TCGATGTACTTCCGAGGAAGCGTGGTGCATAATGCCACGCGCCGCTTGACACTAAAACTC 6275
                |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       3901 TCGATGTACTTCCGAGGAAGCGTGGTGCATAACGCCACGCGCCGCTTGACACTAAAACTC 3960

Qy       6276 GATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGTCACATTAATCGTATATC 6335
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3961 GATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGTCACATTAATCGTATATC 4020

Qy       6336 ACACTACATATTAACAACACTATATCACTTTTATGAGACTCACTATGGGTCTCTAATATA 6395
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4021 ACACTACATATTAACAACACTATATCACTTTTATGAGACTCACTATGGGTCTCTAATATA 4080

Qy       6396 CACTACACATATTTTACTTAAAAACACTATACACACTTTATAAATTCTTTTATAATTTTT 6455
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4081 CACTACACATATTTTACTTAAAAACACTATACACACTTTATAAATTCTTTTATAATTTTT 4140

Qy       6456 CTTTTGTTTTTATTTTGTTTTTAAAATTTC 6485
                ||||||||||||||||||||||||||||||
Db       4141 CTTTTGTTTTTATTTTGTTTTTAAAATTTC 4170


RESULT 3
AF229608
LOCUS       AF229608                4047 bp    RNA     linear   VRL 27-MAR-2000
DEFINITION  Western equine encephalomyelitis virus structural polyprotein gene,
            complete cds.
ACCESSION   AF229608
VERSION     AF229608.1  GI:7330240
KEYWORDS    .
SOURCE      Western equine encephalomyelitis virus
  ORGANISM  Western equine encephalomyelitis virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
            Alphavirus; WEEV complex.
REFERENCE   1  (bases 40 to 816)
  AUTHORS   Uryvaev,L.V., Lebedev,A.Iu., Sokolova,T.M. and Iuferov,V.P.
  TITLE     Primary structure of the nucleocapsid gene C and the protein coded
            by it from the Western equine encephalomyelitis virus
  JOURNAL   Dokl. Akad. Nauk. 344 (3), 397-401 (1995)
  MEDLINE   96097271
   PUBMED   8535272
REFERENCE   2  (bases 40 to 816)
  AUTHORS   Uryvaev,L.V. and Lebedev,A.Iu.
  TITLE     Comparative analysis of primary structure of nucleocapsid protein
            from Western equine encephalomyelitis virus and other alphaviruses
  JOURNAL   Vopr. Virusol. 41 (6), 252-259 (1996)
  MEDLINE   97130212
   PUBMED   8999681
REFERENCE   3  (bases 997 to 2265)
  AUTHORS   Uryvaev,L.V., Lebedev,A.Iu., Sokolova,T.M. and Iuferov,V.P.
  TITLE     Capsid glycoprotein of E2 Encephalitis Virus, Western Equine:
            primary structure of gene and its product
  JOURNAL   Dokl. Akad. Nauk. 357 (1), 134-139 (1997)
  MEDLINE   98100973
   PUBMED   9471275
REFERENCE   4  (bases 1 to 4047)
  AUTHORS   Uryvaev,L.V.
  TITLE     Primary structure of WEEV 26S RNA
  JOURNAL   Unpublished
REFERENCE   5  (bases 1 to 4047)
  AUTHORS   Uryvaev,L.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2000) Molecular Biology, D. I. Ivanovsky
            Institute of Virology, 16 Gamaleya str., Moscow, Russia
FEATURES             Location/Qualifiers
     source          1. .4047
                     /organism="Western equine encephalomyelitis virus"
                     /mol_type="genomic RNA"
                     /strain="McMillan"
                     /db_xref="taxon:11039"
     CDS             40. .3747
                     /codon_start=1
                     /product="structural polyprotein"
                     /protein_id="AAF60166.1"
                     /db_xref="GI:7330241"
                     /translation="MFPYPQLNFPPVYPTNPMAYRDPNPPRCRWRPFRPPLAAQIEDL
                     RRSIANLTFKQRAPNPPPCPPPKKKKSAPEPKPTQPKKKKQQAKKTKRKPKPWKRQRM
                     CMKLESDKTFPIMLNGQVNGYACVVGGRLMKPLHVEGKIDNEQLAAVKLKKASMYDLE
                     YGDVPQNMKSDTLQYTSDKPPGFYNWHHGAVQYENGRFTVPRGVGGKGDSGRPILDNR
                     GRVVAIVLGGANEGTRTALSVVTWNQKGVTIKDTPEGSEPWSLVTALCVLSNVTFPCD
                     KPPVCYSLAPERTLDVLEENVDNPNYDTLLENVLKCPSRRPKRSITDDFTLTSPYLGF
                     CPNCRHSAPCFSPIKIENVWDESDDGSIRIQVSAQFGYNQAGTADVTKFRYMSYDHDH
                     DIEEDSMEKIAISTSGPCRRLGHKGYFLLAQCPPGDSVTVSITSGASENSCTVEFKIR
                     RKFVGREEYLFPPVHGKRVKCHVYDHLKERSAGYITMHRPSPHAYKSYLKEASGEVYI
                     KPPSGKNVTYECKCGDYSTGIVSTPTKMNGCTKAKQCIAYKSDQTKWVFNSPDLIRHT
                     DHSVQGKLHIPFRLIPTVCPVPLAHTPTVTKWFKGITLHLTATRPTLLTTRKLGLRAD
                     ATAEWITGTTSRNFSVGREGLEYVWGNHEPVRVWAQESAPSDPHGWPHEIIIHYYHRH
                     PVYTVIVLCGVALAILVGTASSAACITKARRDCLTPYALAPNATVPTALAVLCCIRPT
                     NAETFGETLNHLWFNNQPFLWAQLCIPLAALIILFRCFSCCMPFLLVAGVCLGKVDAF
                     EHATTVPNVPGIPYKALVERAGYAPLNLEIKVVSSELTPSTNKEYVTCKFHTVIPSPQ
                     VKCCGSLECKASSKADYTCRVFGGVYPFMWGGAQCFCDSENTQLSEAYVEFAPDCTID
                     HAVALRVHTAALKVGLRIVYGNTTAYLDTFVNGVTPGSSRDLKVIAGPISAAFSPFDH
                     KVVIRKGLVYNYDFPEYGAMKPGAFGDIQASSLDATDIVARTDIRLLKPSVKNIHVPY
                     TQAVSGYEMWKNNSGRPLQETAPFGCKIEVEPLRASNCAYGHIPISIDIPDAAFVRSS
                     ESPTILEVSCTADCIYSADFGGSLTLQYKADREGHCPVHSHSTTAVLKEATTHVTATG
                     SITLHFSTSSPQANFIVSLCGKKTTCNAECKPPADHIIGEPHKVDQEFQAAVSKTSWN
                     WLLALFGGASSLIVVGLIVLVCSSMLINTRR"
     mat_peptide     40. .816
                     /product="C-protein"
                     /note="nucleocapsid protein"
                     /citation=[1]
                     /citation=[2]
     mat_peptide     817. .996
                     /product="E3-polypeptide"
     mat_peptide     997. .2265
                     /product="E2-protein"
                     /note="envelope glycoprotein"
                     /citation=[3]
     mat_peptide     2266. .2460
                     /product="6K-polypeptide"
     mat_peptide     2461. .3744
                     /product="E1-protein"
                     /note="envelope glycoprotein"
     3'UTR           3748. .4047
BASE COUNT     1125 a   1041 c    962 g    919 t
ORIGIN

  Query Match           59.4%;  Score 3850.2;  DB 14;  Length 4047;
  Best Local Similarity  97.2%;  Pred. No. 0;
  Matches 3939;  Conservative    0;  Mismatches  108;  Indels    4;  Gaps    2;

Qy       2435 ATAGGTGACGTAGTAGACACGCACCTACCCACCGGCAGAATGTTTCCATACCCTCAGCTG 2494
                ||||||||||||||||||||||||||||||||||| || ||||||||||||||||||||||
Db          1 ATAGGTGACGTAGTAGACACGCACCTACCCACCGCCAAAATGTTTCCATACCCTCAGCTG 60

Qy       2495 AACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACCCTCCTAGG 2554
                ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACCCTCCTAGG 120

Qy       2555 CGCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTAGGAGGTCG 2614
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTAGGAGGTCG 180

Qy       2615 ATAGTCAACTTGACTTTCAAACAACGATCACCTAATCCGCCGCCAGGTCCACCGCCAAAG 2674
                |||| |||||| |||||||||||||||| |||||||||||||||| ||||||||||||||
Db        181 ATAGCCAACTTAACTTTCAAACAACGAGCACCTAATCCGCCGCCATGTCCACCGCCAAAG 240

Qy       2675 AAGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGAAGAAGCAGCAAGCC 2734
                |||||||||||||||| ||||||||||||||||||||||||||||||||| ||||||
Db        241 AAGAAGAAGAGTGCTCCTGAGCCAAAACCTACTCAGCCTAAAAAGAAGAAGCAACAAGCC 300
```

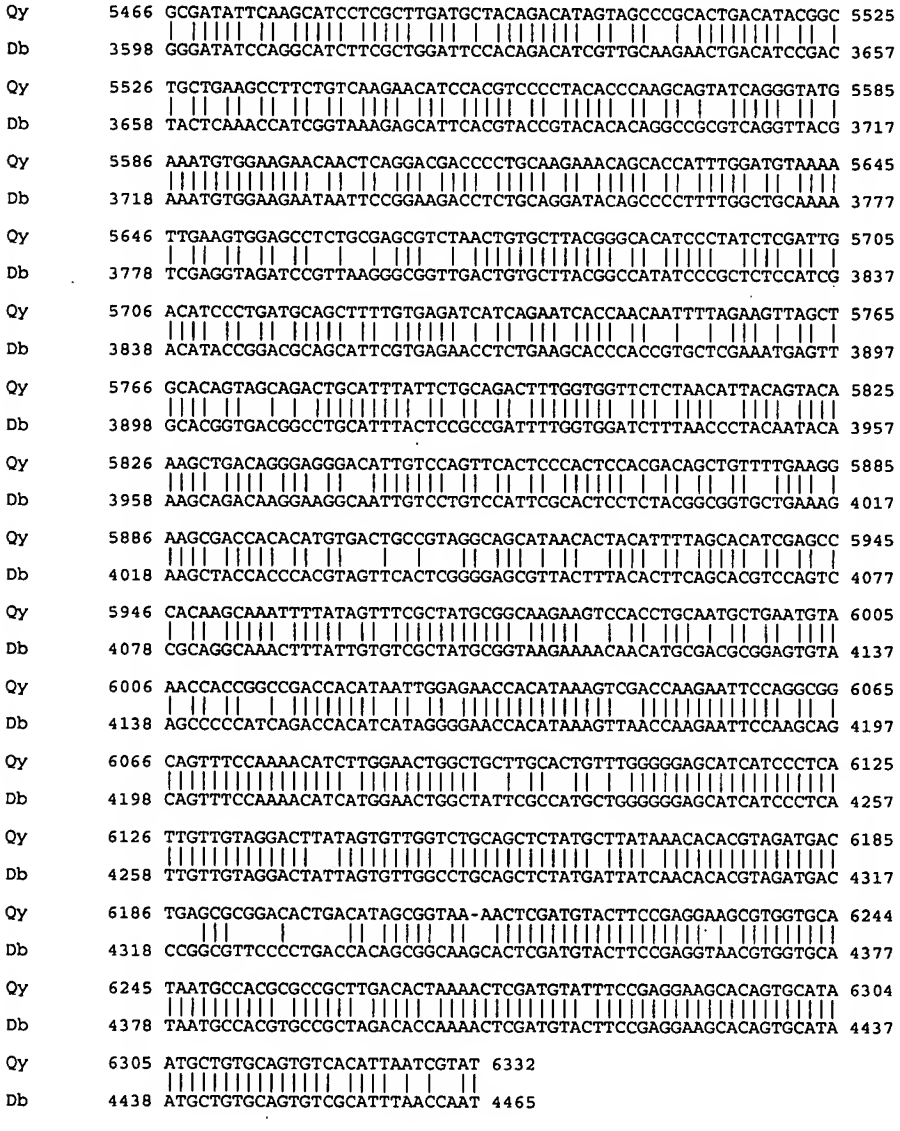

Qy 2735 AAGAGGACGAAACGCAAGCCTAAACCAGGGAAACGACAACGTATGTGTATGAAGTTGGAG 2794
Db 301 AAGAAGACGAAACGCAAGCCTAAACCATGGAAACGACAGCGTATGTGCATGAAGTTGGAG 360
Qy 2795 TCGGACAAGACATTTCCGATCATGCTGAACGGCCAAGTGAATGGATATGCCTGCGTTGTC 2854
Db 361 TCGGACAAGACATTTCCGATCATGCTGAACGGCCAAGTGAATGGATACGCTTGCGTTGTC 420
Qy 2855 GGAGGAAGGCTGATGAAACCACTCCACGTTGAAGGAAAAATTGATAATGAGCAATTAGCG 2914
Db 421 GGAGGAAGGCTGATGAAACCACTCCACGTTGAAGGAAAAATCGATAATGAGCAATTAGCG 480
Qy 2915 GCCGTGAAATTGAAGAAGGCTAGCATGTACGACTTGGAGTACGGCGACGTTCCCCAGAAC 2974
Db 481 GCCGTGAAATTGAAGAAGGCTAGCATGTACGACCTAGAGTATGGCGACGTTCCCCAGAAT 540
Qy 2975 ATGAAATCAGACACGCTGCAGTACACCAGCGACAAACCACCGGGCTTCTACAACTGGCAC 3034
Db 541 ATGAAATCAGACACGTTGCAGTACACCAGCGACAAACCACCGGGCTTCTACAACTGGCAC 600
Qy 3035 CACGGCGCAGTCCAGTATGAGAATGGGAGATTTACCGTACCGAGAGGAGTGGGCGGGAAA 3094
Db 601 CACGGCGCAGTCCAGTATGAGAATGGGAGATTCACCGTACCGCGTGGAGTGGGCGGGAAA 660
Qy 3095 GGCGACAGCGGAAGACCGATCCTGGACAACAGAGGCAGAGTTGTGGCTATTGTTCTAGGA 3154
Db 661 GGCGACAGTGGAAGACCGATCCTAGACAACAGAGGCAGAGTTGTGGCTATTGTTCTAGGA 720
Qy 3155 GGTGCAAATGAGGGCACGCGTACGGCGCTTTCAGTGGTCACTTGGAACCAGAAAGGGGTG 3214
Db 721 GGTGCAAACGAGGGCACGCGTACGGCGCTTTCAGTGGTCACTTGGAACCAAAAAGGGGTA 780
Qy 3215 ACCATTAGGGATACCCCCGAAGGTTCTGAACCGTGGTCACTAGTTACAGCGCTATGCGTG 3274
Db 781 ACCATCAAGGATACCCCCGAAGGTTCTGAACCGTGGTCACTAGTTACAGCGCTGTGCGTG 840
Qy 3275 CTTTCGAATGTCACGTTCCCATGCGACAAACCACCCGTGTGCTATTCACTGACGCCAGAA 3334
Db 841 CTTTCGAATGTCACGTTCCCATGCGACAAACCACCCGTGTGCTATTCACTGGCGCCAGAA 900
Qy 3335 CGAACACTCGACGTGCTCGAAGAGAACGTCGACAATCCAAATTACGACACGCTGCTGGAG 3394
Db 901 CGAACACTCGACGTGCTCGAGGAGAACGTCGACAATCCAAATTACGACACGCTGCTGGAG 960
Qy 3395 AACGTCTTGAAATGTCCATCACGCCGGCCCAAACGAAGCATTACCGATGACTTCACACTG 3454
Db 961 AACGTCTTGAAATGTCCATCACGCCGGCCCAAACGAAGCATTACCGATGACTTCACACTG 1020
Qy 3455 ACCAGTCCCTACCTGGGGTTCTGCCCGTATTGCAGACACTCAACGCCGTGTTTCAGCCCA 3514
Db 1021 ACCAGTCCCTACCTGGGGTTCTGCCCGAATTGCAGACACTCAGCGCCGTGTTTCAGCCCA 1080
Qy 3515 ATAAAAATTGAGAACGTGTGGGACGAATCTGATGATGGATCGATTAGAATCCAGGTCTCG 3574
Db 1081 ATAAAAATAGAGAACGTGTGGGACGAATCTGATGATGGATCGATTAGAATCCAGGTCTCG 1140
Qy 3575 GCACAATTCGGCTACAATCAGGCAGGCACTGCGGATGTCACCAAATTCCGTTACATGTCT 3634
Db 1141 GCACAATTTGGCTACAATCAGGCAGGCACTGCAGATGTCACCAAATTCCGTTACATGTCT 1200
Qy 3635 TTCGACCACGACCATGACATCAAGGAAGACAGTATGGAGAAAATAGCTATCAGCACATCT 3694
Db 1201 TACGACCACGACCATGACATCGAGGAAGACAGTATGGAGAAAATAGCTATTAGCACATCT 1260
Qy 3695 GGACCCTGCCGTCGTCTTGGCCACAAAGGGTACTTCCTGTTAGCTCAATGTCCTCCAGGT 3754
Db 1261 GGACCCTGCCGTCGTCTTGGCCACAAAGGGTACTTCCTGTTAGCTCAATGTCCTCCAGGT 1320
Qy 3755 GACAGTGTAACCGTCAGTATCACGAGCGGAGCATCTGAGAATTCATGCACCGTGGAGAAA 3814
Db 1321 GACAGTGTAACCGTCAGTATCACGAGCGGAGCATCTGAGAATTCATGCACCGTGGAGTTT 1380
Qy 3815 AAGATCAGGAGGAAGTTTGTCGGTAGAGAGGAGTACTTGTTCCCACCCGTCCATGGAAAG 3874
Db 1381 AAGATCAGGAGGAAGTTTGTCGGTAGAGAGGAGTACTTGTTCCCACCTGTCCATGGAAAG 1440
Qy 3875 CTGGTAAAGTGCCACGTTTACGATCACTTGAAGGAGACGTCTGCCGGGTACATAACCATG 3934
Db 1441 AGGGTAAAGTGCCACGTTTACGATCACTTGAAGGAGAGGTCTGCCGGATACATAACCATG 1500
Qy 3935 CACAGGCCAGGCCCACACGCGTATAAGTCCTATCTGGAGGAAGCGTCAGGCGAAGTGTAC 3994
Db 1501 CACAGGCCAAGCCCACACGCGTATAAGTCCTATCTGAAGGAAGCGTCAGGCGAAGTGTAC 1560
Qy 3995 ATTAAACCACCTTCTGGCAAGAACGTCACCTACGAATGTAAGTGTGGCGACTACAGCACA 4054
Db 1561 ATTAAACCACCTTCTGGCAAGAACGTCACCTACGAATGTAAGTGTGGCGACTACAGCACA 1620
Qy 4055 GGTATCGTGAGCACGCGAACGAAGATGAACGGCTGCACTAAAGCAAAACAGTGCATTGCC 4114
Db 1621 GGTATTGTGAGCACGCCAACGAAGATGAACGGCTGCACTAAAGCAAAACAGTGCATTGCC 1680
Qy 4115 TACAAGAGCGACCAAACGAAATGGGTCTTCAACTCGCCGGATCTTATTAGGCACACAGAC 4174
Db 1681 TACAAGAGCGACCAAACGAAATGGGTCTTCAACTCGCCGGATCTTATTAGGCACACAGAC 1740
Qy 4175 CACTCAGTGCAAGGTAAATTGCACATTCCATTCCGCTTGACACCGACAGTCTGCCCGGTT 4234
Db 1741 CACTCAGTGCAAGGTAAACTGCACATTCCATTCCGTTTGATACCGACAGTCTGCCCGGTT 1800
Qy 4235 CCGTTAGCTCACACGCCTACAGTCACGAAGTGGTTCAAAGGCATCACCCTCCACCTGACT 4294
Db 1801 CCGTTAGCTCACACGCCTACAGTCACGAAGTGGTTCAAAGGCATCACCCTCCACCTGACT 1860
Qy 4295 GCAATGCGACCAACATTGCTGACAACGAGAAAATTGGGGCTGCGAGCAGACGCAACAGCA 4354
Db 1861 GCAACGCGACCAACATTGTTGACAACGAGAAAATTGGGGCTGCGAGCAGACGCGACAGCA 1920
Qy 4355 GAATGGATTACAGGGTCTACATCCAGGAATTTTTCTGTGGGGCGAGAAGGGCTGGAGTAC 4414
Db 1921 GAATGGATTACAGGGACTACATCCAGGAATTTTTCTGTGGGGCGAGAAGGGCTGGAGTAC 1980
Qy 4415 GTATGGGGTAACCATGAACCAGTCAGAGTCTGGGCCCAGGAGTCGGCACCAGGCGACCCA 4474
Db 1981 GTATGGGGCAACCATGAACCAGTCAGAGTCTGGGCCCAGGAGTCGGCACCAAGCGACCCA 2040
Qy 4475 CATGGATGGCCGCATGAGATCATCATCCACTATTATCATCGGCATCCAGTCTACACTGTC 4534
Db 2041 CATGGATGGCCGCACGAGATCATCATCCACTATTATCATCGGCATCCAGTCTACACTGTC 2100
Qy 4535 ATTGTGCTGTGTGGTGTCGCTCTTGCTATCCTGGTAGGCACTGCATCATCAGCAGCTTGC 4594
Db 2101 ATTGTGCTGTGCGGTGTCGCTCTTGCTATCCTGGTAGGCACTGCATCGTCAGCAGCTTGT 2160
Qy 4595 ATCGCCAAAGCAAGAAGAGACTGCCTGACGCCATACGCGCTTGCACCGAACGCAACGGTA 4654
Db 2161 ATCACCAAAGCAAGAAGAGACTGCCTGACGCCATACGCGCTTGCACCGAACGCAACGGTA 2220
Qy 4655 CCCACAGCATTAGCGGTTTTGTGCTGCATTCGGCCAACCAACGCTGAAACATTTGGAGAA 4714
Db 2221 CCCACAGCATTAGCAGTTTTGTGCTGTATTCGGCCAACCAATGCAGAAACATTTGGAGAA 2280
Qy 4715 ACTTTGAACCATCTGTGGTTTAACAACCAACCGTTTCTCTGGGCACAGTTGTGCATTCCT 4774
Db 2281 ACTTTGAATCATCTGTGGTTTAACAACCAACCGTTTCTCTGGGCACAGTTGTGCATCCCT 2340
Qy 4775 CTGGCAGCGCTTGTTATTCTGTTCCGCTGCTTTTCATGCTGCATGCCTTTTTTATTGGTT 4834
Db 2341 CTGGCAGCGCTTATTATTCTGTTCCGCTGCTTTTCATGCTGCATGCCTTTTTTATTGGTT 2400
Qy 4835 GCAGGCGTCTGCCTGGGGAAGGTAGACGCCTTCGAACATGCGACCACTGTGCCAAATGTT 4894
Db 2401 GCAGGCGTCTGCCTGGGGAAGGTAGACGCCTTCGAACATGCGACCACTGTGCCAAATGTT 2460
Qy 4895 CCGGGGATCCCGTATAAGGCGTTGGTCGAACGCGCAGGTTACGCGCCACTTAACCTGGAG 4954

```
Db       2461 CCGGGGATCCCGTATAAGGCGTTGGTCGAACGTGCAGGTTACGCGCCACTTAATCTGGAG 2520

Qy       4955 ATCACGGTCGTCTCATCGGAATTAACACCTTCAACTAACAAGGAGTACGTGACCTGCAAA 5014

Db       2521 ATTAAGGTCGTCTCATCGGAATTAACACCCTCAACTAACAAGGAGTACGTGACCTGCAAA 2580

Qy       5015 TTCCACACAGTCATTCCTTCACCACAAGTTAAATGCTGCGGGTCCCTCGAGTGCAAGGCA 5074

Db       2581 TTCCACACAGTCATTCCTTCACCACAAGTTAAATGCTGCGGGTCCCTCGAGTGCAAGGCA 2640

Qy       5075 TCCTCAAAGGCGGATTACACATGCCGCGTTTTTGGCGGTGTGTACCCTTTCATGTGGGGA 5134

Db       2641 TCCTCAAAAGCGGATTACACATGCCGCGTTTTTGGCGGTGTGTACCCTTTCATGTGGGGA 2700

Qy       5135 GGCGCACAATGCTTCTGTGACAGTGAGAACACACAACTGAGTGAGGCGTACGTCGAGTTC 5194

Db       2701 GGCGCACAGTGCTTCTGTGACAGTGAGAACACACAACTGAGTGAGGCATACGTCGAGTTC 2760

Qy       5195 GCTCCAGACTGCACTATAGATCACGCAGTCGCACTAAAAGTTCACACAGCTGCTCTGAAA 5254

Db       2761 GCTCCAGACTGCACTATAGATCATGCAGTCGCACTAAGAGTTCACACAGCTGCTCTGAAA 2820

Qy       5255 GTCGGCCTGCGTATAGTATACGGCAACACCACCGCGCACCTGGATACGTTTGTCAATGGC 5314

Db       2821 GTCGGCCTGCGTATAGTATACGGCAATACCACCGCGTACCTGGATACGTTCGTCAATGGC 2880

Qy       5315 GTCACGCCAGGTTCCTCACGGGACCTGAAGGTCATAGCAGGGCCGATATCAGCCGCTTTT 5374

Db       2881 GTCACGCCAGGTTCCTCACGGGACCTGAAGGTCATAGCAGGGCCGATATCAGCTGCTTTT 2940

Qy       5375 TCACCCTTTGACCATAAGGTCGTCATCAGAAAGGGGCTTGTTTACAACTACGACTTCCCT 5434

Db       2941 TCACCCTTTGACCATAAGGTCGTCATCAGAAAGGGGCTTGTTTACAACTACGACTTCCCT 3000

Qy       5435 GAGTATGGAGCTATGAAACCAGGAGCGTTCGGCGATATTCAAGCATCCTCGCTTGATGCT 5494

Db       3001 GAGTATGGAGCTATGAAACCAGGAGCGTTCGGCGATATTCAAGCATCCTCGCTTGATGCC 3060

Qy       5495 ACAGACATAGTAGCCCGCACTGACATACGGCTGCTGAAGCCTTCTGTCAAGAACATCCAC 5554

Db       3061 ACAGACATAGTAGCCCGCACTGACATACGGCTGCTGAAGCCTTCTGTCAAGAACATCCAC 3120

Qy       5555 GTCCCCTACACCCAAGCAGTATCAGGGTATGAAATGTGGAAGAACAACTCAGGACGACCC 5614

Db       3121 GTCCCCTACACCCAAGCAGTATCAGGGTATGAAATGTGGAAGAACAACTCAGGACGACCC 3180

Qy       5615 CTGCAAGAAACAGCACCATTTGGATGTAAAATTGAAGTGGAGCCTCTGCGAGCGTCTAAC 5674

Db       3181 CTGCAAGAAACAGCACCATTTGGATGTAAAATTGAAGTGGAGCCTCTGCGAGCGTCTAAC 3240

Qy       5675 TGTGCTTACGGGCACATCCCTATCTCGATTGACATCCCTGATGCAGCTTTTGTGAGATCA 5734

Db       3241 TGTGCTTATGGGCACATCCCTATCTCGATTGACATCCCTGATGCAGCTTTTGTGAGATCA 3300

Qy       5735 TCAGAATCACCAACAATTTTAGAAGTTAGCTGCACAGTAGCAGACTGCATTTATTCTGCA 5794

Db       3301 TCAGAATCACCAACAATTTTAGAAGTCAGCTGCAC---AGCAGACTGCATTTATTCTGCA 3357

Qy       5795 GACTTTGGTGGTTCTCTAACATTACAGTACAAAGCTGACAGGGAGGGACATTGTCCAGTT 5854

Db       3358 GACTTTGGTGGTTCGCTAACACTACAGTACAAAGCTGACAGGGAGGGACATTGTCCAGTT 3417

Qy       5855 CACTCCCACTCCACGACAGCTGTTTTGAAGGAAGCGACCACACATGTGACTGCCGTAGGC 5914

Db       3418 CACTCCCACTCCACTACAGCTGTTTTGAAGGAAGCGACCACACATGTGACTGCCACAGGC 3477

Qy       5915 AGCATAACACTACATTTTAGCACATCGAGCCCACAAGCAAATTTTATAGTTTCGCTATGC 5974

Db       3478 AGCATAACACTACATTTTAGCACATCGAGCCCACAAGCAAATTTCATAGTTTCGCTATGC 3537

Qy       5975 GGCAAGAAGTCCACCTGCAATGCTGAATGTAAACCACCGGCCGACCACATAATTGGAGAA 6034

Db       3538 GGCAAGAAGACCACCTGCAATGCTGAATGTAAACCACCGGCCGACCACATAATTGGAGAA 3597

Qy       6035 CCACATAAAGTCGACCAAGAATTCCAGGCGGCAGTTTCCAAAACATCTTGGAACTGGCTG 6094

Db       3598 CCACATAAAGTCGACCAAGAATTCCAGGCGGCAGTTTCCAAAACATCTTGGAATTGGCTG 3657

Qy       6095 CTTGCACTGTTTGGGGGAGCATCATCCCTCATTGTTGTAGGACTTATAGTGTTGGTCTGC 6154

Db       3658 CTTGCACTGTTTGGGGGAGCATCATCCCTCATTGTTGTAGGACTTATAGTGTTGGTCTGC 3717

Qy       6155 AGCTCTATGCTTATAAACACACGTAGATGACTGAGCGCGGACACTGACATAGCGGTAAAA 6214

Db       3718 AGCTCTATGCTCATAAACACACGTAGATGACTGGGCGCGGCCACTGACATAGCGGTAAAA 3777

Qy       6215 CTCGATGTACTTCCGAGGAAGCGTGGTGCATAATGCCACGCGCCGCTTGACACTAAAACT 6274

Db       3778 CTCGATGTACTTCCGAGGAAGCGTGGTGCATAATGCCACGCCCCGCTTGACATTAAAACT 3837

Qy       6275 CGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGTCACATTAATCGTATAT 6334

Db       3838 CGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGTCACATTAATCGTATAT 3897

Qy       6335 CACACTACATATTAACAACACTATATCACTTTTATGAGACTCACTATGGGTCTCTAATAT 6394

Db       3898 CACACTACATATTAACAACACTATATCATTTTTATGAGACTCA-TATGGGTCTCTAATAT 3956

Qy       6395 ACACTACACATATTTTACTTAAAAACACTATACACACTTTATAAATTCTTTTATAATTTT 6454

Db       3957 ACACTATACATATTTTACTTAAAAACACTATACACACTTTATAAATTCTTTTATAATTTT 4016

Qy       6455 TCTTTTGTTTTTATTTTGTTTTTAAAATTTC 6485

Db       4017 TCTTTTGTTTTTATTTTGTTTTTAAAATTTC 4047
```

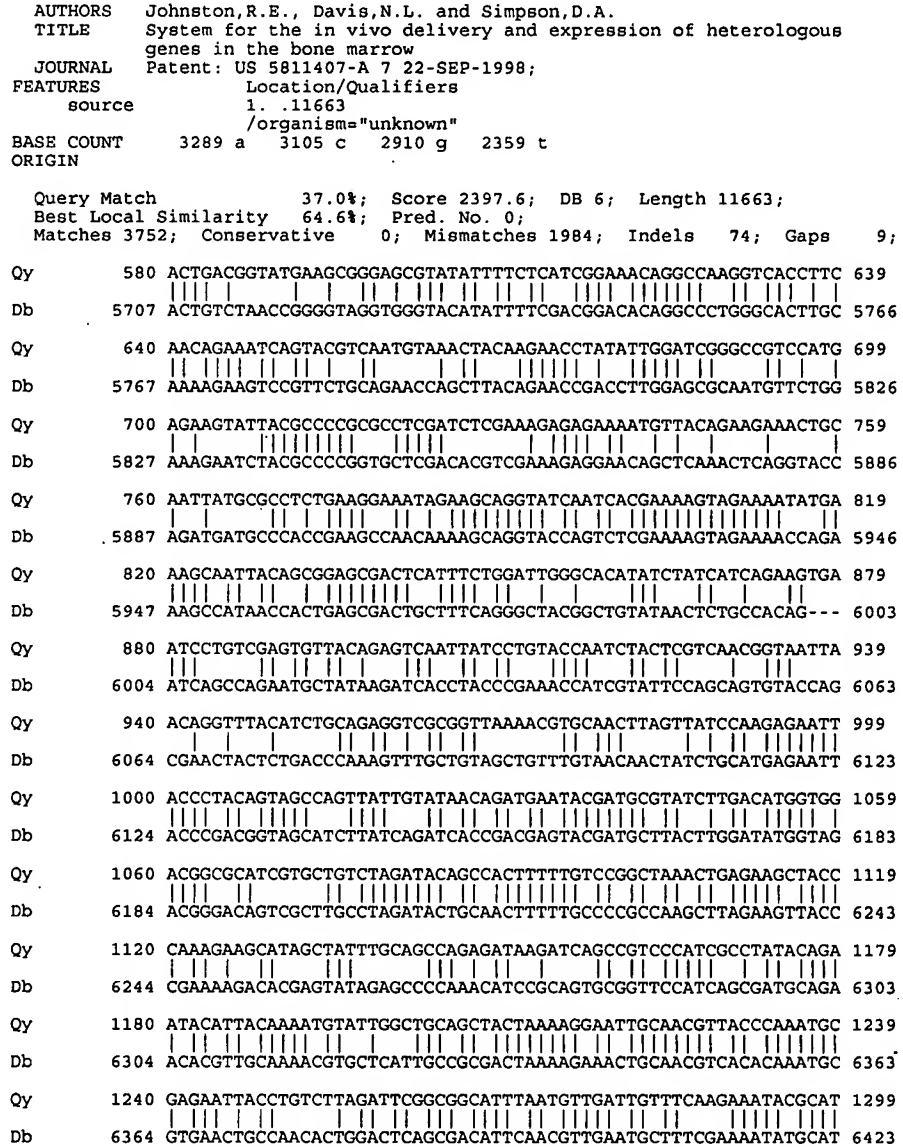
```
RESULT 4
AF339476
LOCUS       AF339476               4692 bp ss-RNA     linear   VRL 02-FEB-2003
DEFINITION  Highlands J virus strain B-230 nonstructural protein 4 gene,
            partial cds; and structural polyprotein precursor, gene, complete
            cds.
ACCESSION   AF339476
VERSION     AF339476.1  GI:28193935
KEYWORDS    .
SOURCE      Highlands J virus
  ORGANISM  Highlands J virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
            Alphavirus; WEEV complex.
REFERENCE   1  (bases 1 to 4692)
  AUTHORS   Kinney,R.M. and Pfeffer,M.
  TITLE     Nucleotide sequence analyses of the 26S mRNAs of viruses of the
            genus Alphavirus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4692)
  AUTHORS   Kinney,R.M. and Pfeffer,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JAN-2001) Arbovirus Diseases Branch, Centers for
            Disease Control and Prevention, P.O. Box 2087, Fort Collins, CO
            80522, USA
FEATURES             Location/Qualifiers
     source          1. .4692
                     /organism="Highlands J virus"
                     /virion
                     /mol_type="genomic RNA"
                     /strain="B-230"
                     /db_xref="taxon:11024"
                     /country="USA"
                     /note="isolated in 1960; virus passage P5/SM3/BHK21-1"
     CDS             <1. .564
                     /codon_start=1
                     /product="nonstructural protein 4"
                     /protein_id="AAO33322.1"
                     /db_xref="GI:28193936"
```

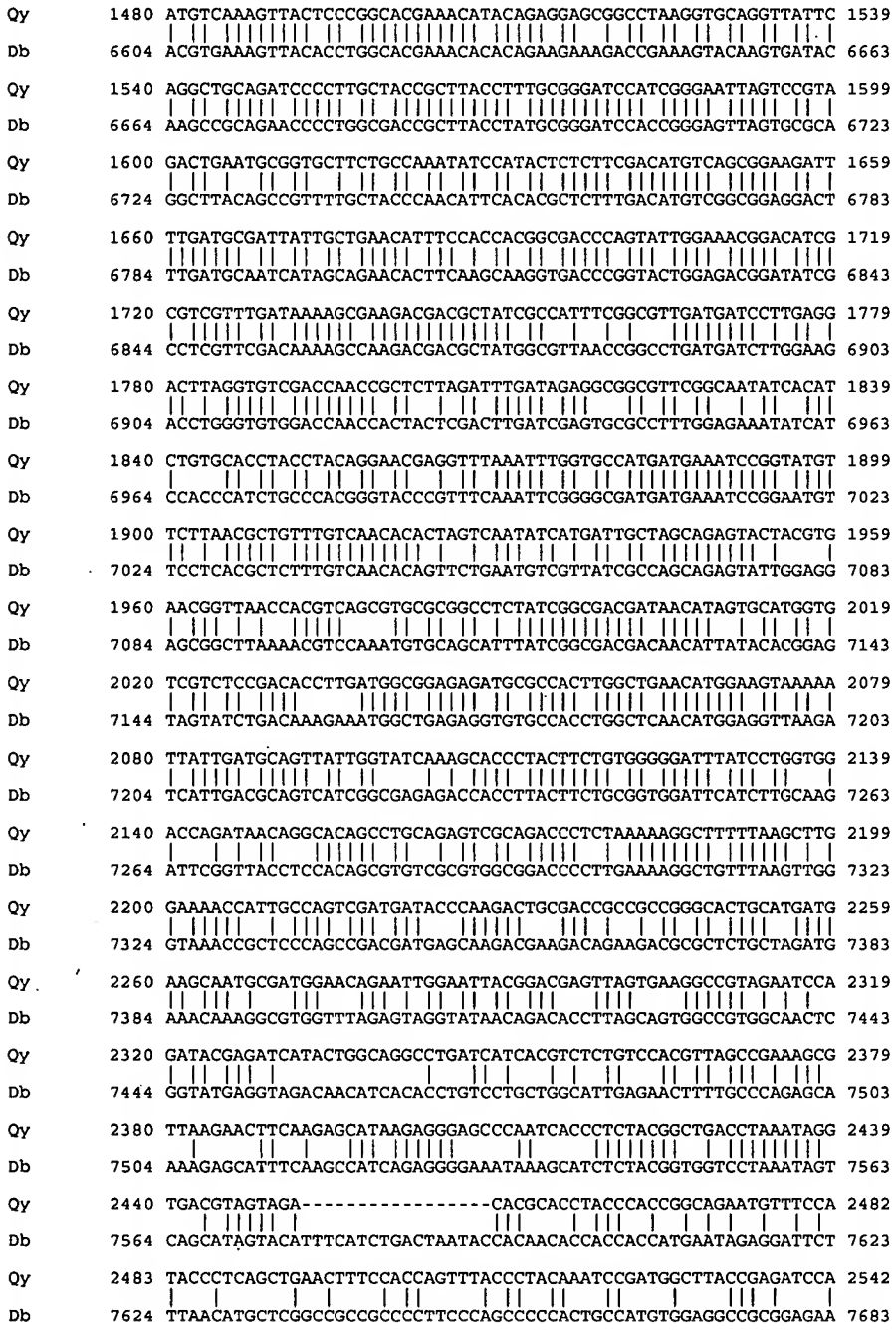
```
                     /translation="FKFGAMMKSGMFLTLFVNTLVNIMIASRVLRERLTTSACAAFIG
                     DDNIVHGVVSDNLMADRCATWLNMEVKIIDAVIGVKAPYFCGGFILVDQITGTACRVA
                     DPLKRLFKLGKPLPLDDDQDVDRRRALHDESLRWNRIGITSELVKAVESRYEVVLSSL
                     IIIALSTLADSVKNFKRIRGNPITLYG"
     mRNA            565. .4692
                     /product="structural polyprotein precursor"
                     /note="26S"
     5'UTR           565. .605
     CDS             606. .4316
                     /codon_start=1
                     /product="structural polyprotein precursor"
                     /protein_id="AAO33323.1"
                     /db_xref="GI:28193937"
                     /translation="MFPYPQLTFPPMYQPNPMAYRDPNPPRRRWRPFRPPLAAQIEDL
                     RRSIANLTLKQRAPNPPAGPPAKKKKTQPKPKATPPKKKKQQVKKQKRKPKPGKRQRM
                     CMKLESDKTFPILLNGQVNGYACVVGGRLMKPLHVEGKIDNEQLAAIKLKKASMYDLE
                     YGDVPQNMKSDTLQYTSEKPPGFYNWHHGAVQYENGRFTVPRGVGGKGDSGRPILDNK
                     GRVVAIVLGGANEGSRTALSVVTWNQKGVTIKDTPEGTEQWSLITAMCVLANVTFPCD
                     KPPVCYSLAPERTLDVLEENVDNPGYDTLLENVLKCPSRRQKRSITDDFTLTSPYLGH
                     CPYCLHATPCFSPIKIENVWDESDDGTIRIQVSAQLGYNQAGTADPTKFRYMSYEQDH
                     DIKEASMDKIAISTSGPCSRLGHKGYFLLARCPPGDSVTVSITSGASATSCTVEKKIR
                     RKFVGREEYLLPPVHGKLIKCHVYDHLKETTAGYITMHRPGPHAYATYVEESSGKVYI
                     RPPSGKNVTYECKCGDYSTGTVNTRTKMPGCTKKKQCIAYKHDQTKWVFNSPDLIRHS
                     DHAVQGKLHIPFKLTATACPVPLAHTPTVEKWFKGVTLHLTASHPTLLTTRKLGPRAE
                     PTSEWIVGTVSRNFSVGREGLEYTWGNHDPVRVWSQESAPGDPHGWPHEIIVHYYHRH
                     PLYTIAVLCGLVLITVIGIASAAACISKARRDCLTPYALAPNAAVPTLLAVLCCIRPT
                     HAETLGESLGHLWLNNQPLLWAQLCLPLAALIILFRFFSCCLPFLLVAGVCLGKADAY
                     EHATTVPNVPGVPYKALVERAGYAPLNLEVTVVSSELIPSTNKEYVTCKFHTIIPSPQ
                     VKCCGSLECQASRKADYTCRVFGGVYPFMWGGAQCFCDSENTQLSEAYVEFAPDCTAD
                     HAVALKVHTAALKVGLQIVYGNTSTRLDTFVNGVTPGSSRDLKVIAGPISAAFTPFDH
                     KVVIRKGKVYNYDFPEYGAMKPGVFGDIQASSLDSTDIVARTDIRLLKPSVKSIHVPY
                     TQAASGYEMWKNNSGRPLQDTAPFGCKIEVDPLRAVDCAYGHIPLSIDIPDAAFVRTS
                     EAPTVLEMSCTVTACIYSADFGGSLTLQYKADKEGNCPVHSHSSTAVLKEATTHVVHS
                     GSVTLHFSTSSPQANFIVSLCGKKTTCDAECKPPSDHIIGEPHKVNQEFQAAVSKTSW
                     NWLFAMLGGASSLIVVGLLVLACSSMIINTRR"
     mat_peptide     606. .1382
                     /product="capsid protein"
     mat_peptide     1383. .1442
                     /product="E3 protein"
     mat_peptide     1443. .2831
                     /product="E2 envelope glycoprotein"
     mat_peptide     2832. .2996
                     /product="6K protein"
     mat_peptide     2997. .4313
                     /product="E1 envelope glycoprotein"
     3'UTR           4317. .4692
BASE COUNT     1324 a   1183 c   1105 g   1080 t
ORIGIN

  Query Match           41.7%;  Score 2705.6;  DB 14;  Length 4692;
  Best Local Similarity  75.6%;  Pred. No. 0;
  Matches 3380;  Conservative    0;  Mismatches 1084;  Indels    4;  Gaps    2;

Qy       1866 TTTAAATTTGGTGCCATGATGAAATCCGGTATGTTCTTAACGCTGTTTGTCAACACACTA 1925
Db          1 TTTAAGTTCGGAGCCATGATGAAATCAGGGATGTTTTTGACCTTGTTTGTTAATACTCTA 60

Qy       1926 GTCAATATCATGATTGCTAGCAGAGTACTACGTGAACGGTTAACCACGTCAGCGTGCGCG 1985
Db         61 GTTAACATCATGATAGCCAGTCGGGTACTTAGAGAAAGGTTGACCACTTCAGCGTGTGCA 120

Qy       1986 GCCTCTATCGGCGACGATAACATAGTGCATGGTGTCGTCTCCGACACCTTGATGGCGGAG 2045
Db        121 GCCTTTATCGGAGACGACAATATAGTACACGGTGTAGTATCCGATAATTTGATGGCTGAT 180

Qy       2046 AGATGCGCCACTTGGCTGAACATGGAAGTAAAAATTATTGATGCAGTTATTGGTATCAAA 2105
Db        181 AGGTGTGCCACGTGGCTAAATATGGAAGTAAAAATCATTGATGCCGTGATCGGCGTAAAA 240

Qy       2106 GCACCCTACTTCTGTGGGGGATTTATCCTGGTGGACCAGATAACAGGCACAGCCTGCAGA 2165
Db        241 GCACCATACTTTTGCGGAGGGTTCATCCTGGTCGATCAAATTACAGGTACAGCGTGTAGG 300

Qy       2166 GTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTGGAAAACCATTGCCAGTCGATGATACC 2225
Db         301 GTAGCAGACCCCCTCAAGCGTTTGTTTAAACTGGGTAAACCACTCCCTTTAGACGATGAT 360

Qy       2226 CAAGACTGCGACCGCCGCCGGGCACTGCATGATGAAGCAATGCGATGGAACAGAATTGGA 2285
Db         361 CAGGATGTAGACCGCCGTCGAGCGTTGCACGATGAGTCTCTAAGGTGGAATAGGATAGGT 420

Qy       2286 ATTACGGACGAGTTAGTGAAGGCCGTAGAATCCAGATACGAGATCATACTGGCAGGCCTG 2345
Db         421 ATAACCTCTGAGCTTGTTAAGGCGGTTGAATCCCGATATGAGGTTGTCCTGTCTAGCTTA 480

Qy       2346 ATCATCACGTCTCTGTCCACGTTAGCCGAAAGCGTTAAGAACTTCAAGAGCATAAGAGGG 2405
Db         481 ATTATTATTGCGCTGTCAACACTTGCAGATAGCGTGAAGAACTTTAAGCGGATAAGAGGA 540

Qy       2406 AGCCCAATCACCCTCTACGGCTGACCTAAATAGGTGACGTAGTAGACACGCACCTACCCA 2465
Db         541 AACCCCATTACTCTTTACGGCTGACCTTAATAGGTGACGTAGTAGATACGCACCTA---A 597

Qy       2466 CCGGCAGAATGTTTCCATACCCTCAGCTGAACTTTCCACCAGTTTACCCTACAAATCCGA 2525
Db         598 CCGCGAAAATGTTTCCTTATCCGCAATTGACTTTTCCACCTATGTATCAACCCAATCCTA 657

Qy       2526 TGGCTTACCGAGATCCAAACCCTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCCGCTGG 2585
Db         658 TGGCATACAGAGACCCGAACCCTCCTAGGCGCAGGTGGAGGCCATTTAGGCCCCCACTTG 717

Qy       2586 CTGCTCAAATCGAAGATCTTAGGAGGTCGATAGTCAACTTGACTTTCAAACAACGATCAC 2645
Db         718 CAGCACAAATAGAGGACTTGAGGAGGTCCATCGCAAATTTGACGCTTAAACAGAGAGCAC 777

Qy       2646 CTAATCCGCCGCCAGGTCCACCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAAACCTA 2705
Db         778 CCAATCCACCTGCCGGTCCTCCAGCTAAGAAGAAGAAGACTCAACCGAAACCGAAGGCCA 837

Qy       2706 CTCAGCCTAAAAAGAAGAAGCAGCAAGCCAAGAGGACGAAACGCAAGCCTAAACCAGGGA 2765
Db         838 CGCCACCTAAGAAGAAGAAGCAGCAGGTTAAGAAACAAAAACGTAAACCGAAACCTGGAA 897

Qy       2766 AACGACAACGTATGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATCATGCTGAACG 2825
Db         898 AAAGGCAACGCATGTGTATGAAATTGGAATCTGACAAAACCTTCCCGATTCTGTTGAACG 957

Qy       2826 GCCAAGTGAATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACTCCACGTTG 2885
Db         958 GGCAAGTAAATGGCTACGCATGCGTTGTCGGCGGAAGGCTAATGAAACCATTGCATGTCG 1017

Qy       2886 AAGGAAAAATTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAGCATGTACG 2945
Db        1018 AAGGTAAAATTGATAACGAACAGTTGGCAGCCATCAAACTGAAGAAAGCCAGTATGTACG 1077

Qy       2946 ACTTGGAGTACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTACACCAGCG 3005
Db        1078 ATCTGGAGTACGGTGACGTACCCCAGAATATGAAATCAGACACTCTGCAATACACCAGCG 1137

Qy       3006 ACAAACCACCGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAATGGGAGAT 3065
Db        1138 AAAAACCACCTGGATTCTATAACTGGCATCATGGCGCAGTCCAGTATGAAAATGGCAGAT 1197

Qy       3066 TTACCGTACCGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGACAACA 3125
Db        1198 TCACCGTTCCACGCGGAGTCGGTGGAAAAGGAGACAGTGGGAGACCCATCCTGGACAACA 1257

Qy       3126 GAGGCAGAGTTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGCGCTTT 3185
Db        1258 AAGGCCGGGTAGTGGCTATTGTCCTGGGAGGGGCGAACGAAGGTTCAAGGACGGCTCTGT 1317

Qy       3186 CAGTGGTCACTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTCTGAAC 3245
Db        1318 CGGTGGTTACATGGAACCAAAAAGGGGTGACCATCAAGGATACCCCTGAAGGCACGGAAC 1377

Qy       3246 CGTGGTCACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGACAAAC 3305
```

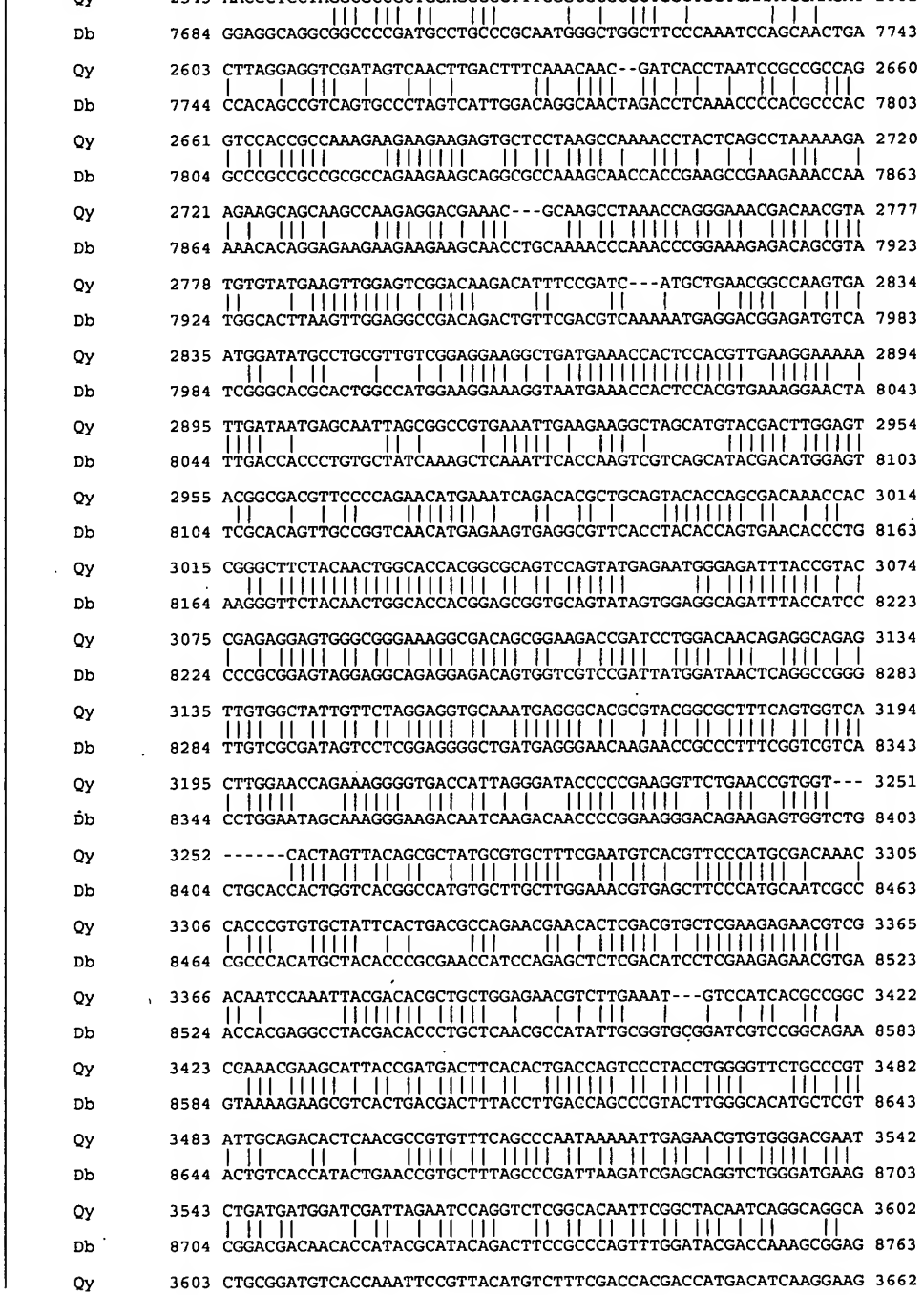

Db 1378 AATGGTCATTGATCACTGCAATGTGCGTCCTTGCAAATGTCACTTTCCCGTGTGACAAAC 1437
Qy 3306 CACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAACGTCG 3365
Db 1438 CTCCTGTCTGCTATTCGTTAGCACCAGAACGCACACTTGATGTGCTGGAAGAAAACGTGG 1497
Qy 3366 ACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAATGTCCATCACGCCGGCCCA 3425
Db 1498 ACAATCCAGGTTACGACACCCTGCTAGAAAATGTACTGAAGTGCCCCTCACGTCGACAGA 1557
Qy 3426 AACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCCCGTATT 3485
Db 1558 AACGGAGCATTACGGACGATTTCACCTTGACAAGCCCGTACTTAGGCCACTGTCCGTATT 1617
Qy 3486 GCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACGAATCTG 3545
Db 1618 GCCTCCACGCTACGCCCTGCTTTAGTCCCATTAAAATTGAAAACGTTTGGGACGAGTCAG 1677
Qy 3546 ATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAGGCACTG 3605
Db 1678 ACGACGGTACAATACGAATACAAGTTTCCGCCCAATTAGGTTACAACCAGGCAGGTACGG 1737
Qy 3606 CGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAAGGAAGACA 3665
Db 1738 CAGACCCAACTAAGTTCAGGTACATGTCCTACGAGCAAGACCATGACATCAAAGAAGCTA 1797
Qy 3666 GTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCACAAAGGGT 3725
Db 1798 GTATGGACAAGATCGCTATCAGTACATCTGGGCCGTGCAGCCGCTTGGGGCATAAAGGCT 1857
Qy 3726 ACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCACGAGCGGAG 3785
Db 1858 ATTTTCTGCTGGCCAGATGCCCACCAGGAGACAGCGTAACCGTCAGCATCACCAGTGGAG 1917
Qy 3786 CATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTAGAGAGG 3845
Db 1918 CATCAGCGACATCCTGCACGGTCGAGAAGAAGATAAGGAGGAAGTTCGTCGGCAGAGAGG 1977
Qy 3846 AGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATCACTTGA 3905
Db 1978 AGTACCTGCTCCCACCAGTGCACGGCAAGCTAATCAAGTGTCATGTATATGACCACCTAA 2037
Qy 3906 AGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAAGTCCT 3965
Db 2038 AGGAGACAACAGCAGGATATATAACTATGCACAGACCAGGGCCACATGCGTACGCAACCT 2097
Qy 3966 ATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGTCACCT 4025
Db 2098 ACGTGGAAGAATCATCCGGCAAGGTGTACATCCGTCCACCTTCCGGAAAGAACGTGACGT 2157
Qy 4026 ACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGATGAACG 4085
Db 2158 ACGAATGCAAATGTGGAGATTACAGCACCGGCACAGTTAACACTAGAACTAAGATGCCAG 2217
Qy 4086 GCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGTCTTCA 4145
Db 2218 GATGCACGAAGAAGAAACAGTGCATAGCTTATAAACACGACCAGACCAAATGGGTCTTTA 2277
Qy 4146 ACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTCCAT 4205
Db 2278 ATTCACCAGACCTAATTCGTCACTCAGACCATGCAGTGCAAGGAAAACTGCACATTCCCT 2337
Qy 4206 TCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGAAGT 4265
Db 2338 TTAAACTAACTGCTACGGCATGCCCAGTGCCACTCGCTCACACACCAACGGTTGAGAAAT 2397
Qy 4266 GGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGAGAA 4325
Db 2398 GGTTCAAAGGCGTCACGCTGCACCTGACTGCTAGTCATCCCACACTGTTGACTACGCGTA 2457
Qy 4326 AATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGAATT 4385
Db 2458 AGCTAGGACCGCGAGCTGAACCCACGTCTGAATGGATCGTTGGGACCGTGTCCAGAAACT 2517
Qy 4386 TTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAGTCT 4445
Db 2518 TCTCCGTAGGGAGGGAGGGTCTCGAATATACTTGGGGAAATCACGACCCCGTGCGAGTGT 2577
Qy 4446 GGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCCACT 4505
Db 2578 GGTCTCAAGAGTCTGCACCTGGGGACCCTCACGGTTGGCCGCACGAGATTATCGTACATT 2637
Qy 4506 ATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTATCC 4565
Db 2638 ACTACCATCGGCATCCGCTATACACCATTGCCGTGCTGTGCGGACTCGTGCTAATAACAG 2697
Qy 4566 TGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGACGC 4625
Db 2698 TGATCGGCATCGCTTCTGCAGCGGCCTGCATATCGAAAGCGCGCAGAGATTGCCTGACAC 2757
Qy 4626 CATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCATTC 4685
Db 2758 CGTACGCGTTGGCACCCAACGCTGCAGTTCCAACGTTGCTGGCAGTCTTATGCTGCATAA 2817
Qy 4686 GGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACCAAC 4745
Db 2818 GACCTACGCACGCCGAGACGCTAGGCGAATCACTTGGACACCTCTGGCTTAACAATCAAC 2877
Qy 4746 CGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCTGCT 4805
Db 2878 CTCTTTTATGGGCGCAGCTGTGCCTTCCACTTGCCGCACTTATCATCCTGTTCAGGTTCT 2937
Qy 4806 TTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACGCCT 4865
Db 2938 TCTCATGCTGCCTCCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGAAAGGCAGACGCCT 2997
Qy 4866 TCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCGAAC 4925
Db 2998 ACGAACATGCGACCACTGTGCCAAATGTTCCGGGAGTCCCGTATAAAGCGCTAGTGGAAA 3057
Qy 4926 GCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACACCTT 4985
Db 3058 GGGCCGGGTATGCACCCCTCAACCTAGAAGTCACCGTTGTGTCGTCTGAACTGATACCGT 3117
Qy 4986 CAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAGTTA 5045
Db 3118 CAACAAATAAAGAATATGTGACTTGCAAATTCCATACAATCATTCCATCACCCCAGGTGA 3177
Qy 5046 AATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCGTTT 5105
Db 3178 AGTGTTGCGGTTCCCTTGAATGTCAAGCCTCACGCAAGGCAGACTATACCTGTCGCGTGT 3237
Qy 5106 TTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGAACA 5165
Db 3238 TCGGAGGTGTCTACCCATTTATGTGGGGAGGCGCCCAGTGCTTCTGCGATAGCGAAAACA 3297
Qy 5166 CACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAGTCG 5225
Db 3298 CGCAACTGAGCGAAGCGTACGTTGAATTCGCTCCAGACTGCACAGCAGATCATGCAGTCG 3357
Qy 5226 CACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACACCA 5285
Db 3358 CACTGAAAGTCCACACAGCTGCCTTGAAGGTCGGCCTGCAAATAGTATATGGGAACACAT 3417
Qy 5286 CCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGAAGG 5345
Db 3418 CAACACGCCTGGATACATTTGTCAATGGGGTAACACCAGGCTCATCACGCGACCTGAAAG 3477
Qy 5346 TCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCAGAA 5405
Db 3478 TGATAGCAGGCCCTATCTCAGCAGCGTTCACACCTTTCGACCACAAAGTGGTCATTCGAA 3537
Qy 5406 AGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGTTCG 5465
Db 3538 AAGGTAAAGTGTACAACTACGACTTCCCAGAGTACGGTGCTATGAAGCCAGGCGTGTTTG 3597

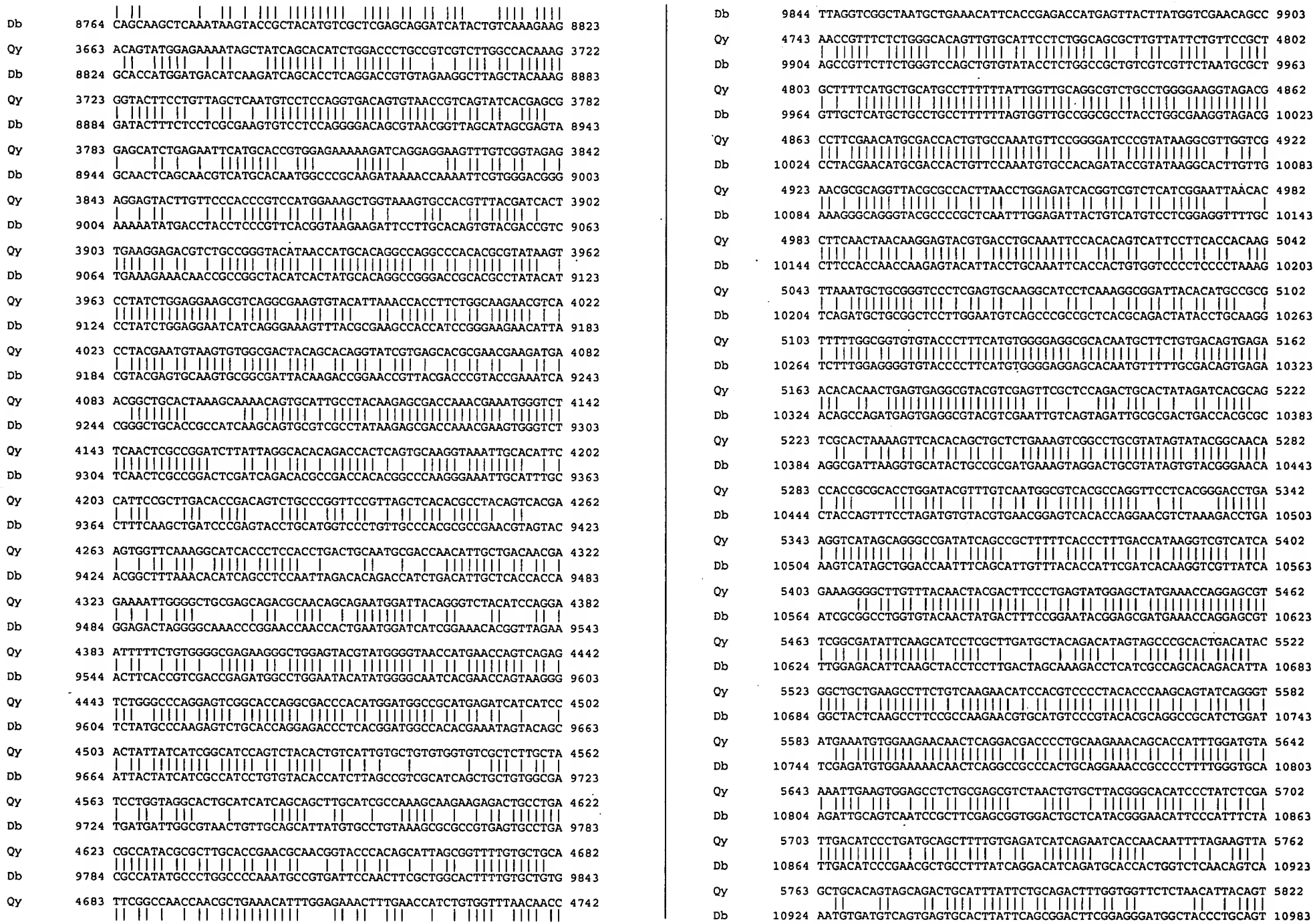


```
RESULT 5
AR042411
LOCUS       AR042411               11663 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 7 from patent US 5811407.
ACCESSION   AR042411
VERSION     AR042411.1  GI:5962907
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 11663)
  AUTHORS   Johnston,R.E., Davis,N.L. and Simpson,D.A.
  TITLE     System for the in vivo delivery and expression of heterologous
            genes in the bone marrow
  JOURNAL   Patent: US 5811407-A 7 22-SEP-1998;
FEATURES             Location/Qualifiers
     source          1. .11663
                     /organism="unknown"
BASE COUNT     3289 a   3105 c   2910 g   2359 t
ORIGIN

  Query Match             37.0%;  Score 2397.6;  DB 6;  Length 11663;
  Best Local Similarity   64.6%;  Pred. No. 0;
  Matches 3752;  Conservative    0;  Mismatches 1984;  Indels   74;  Gaps    9;
```







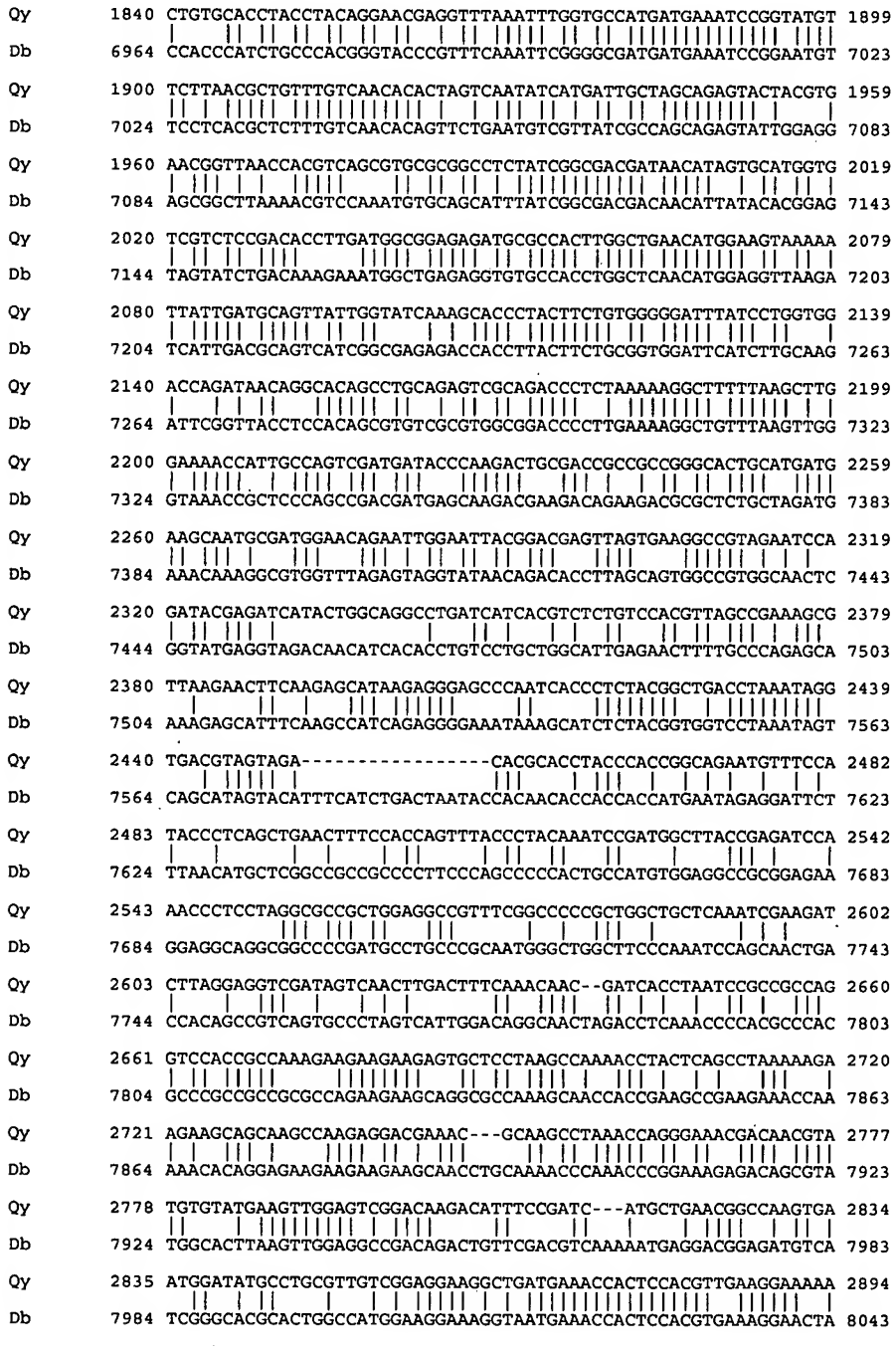


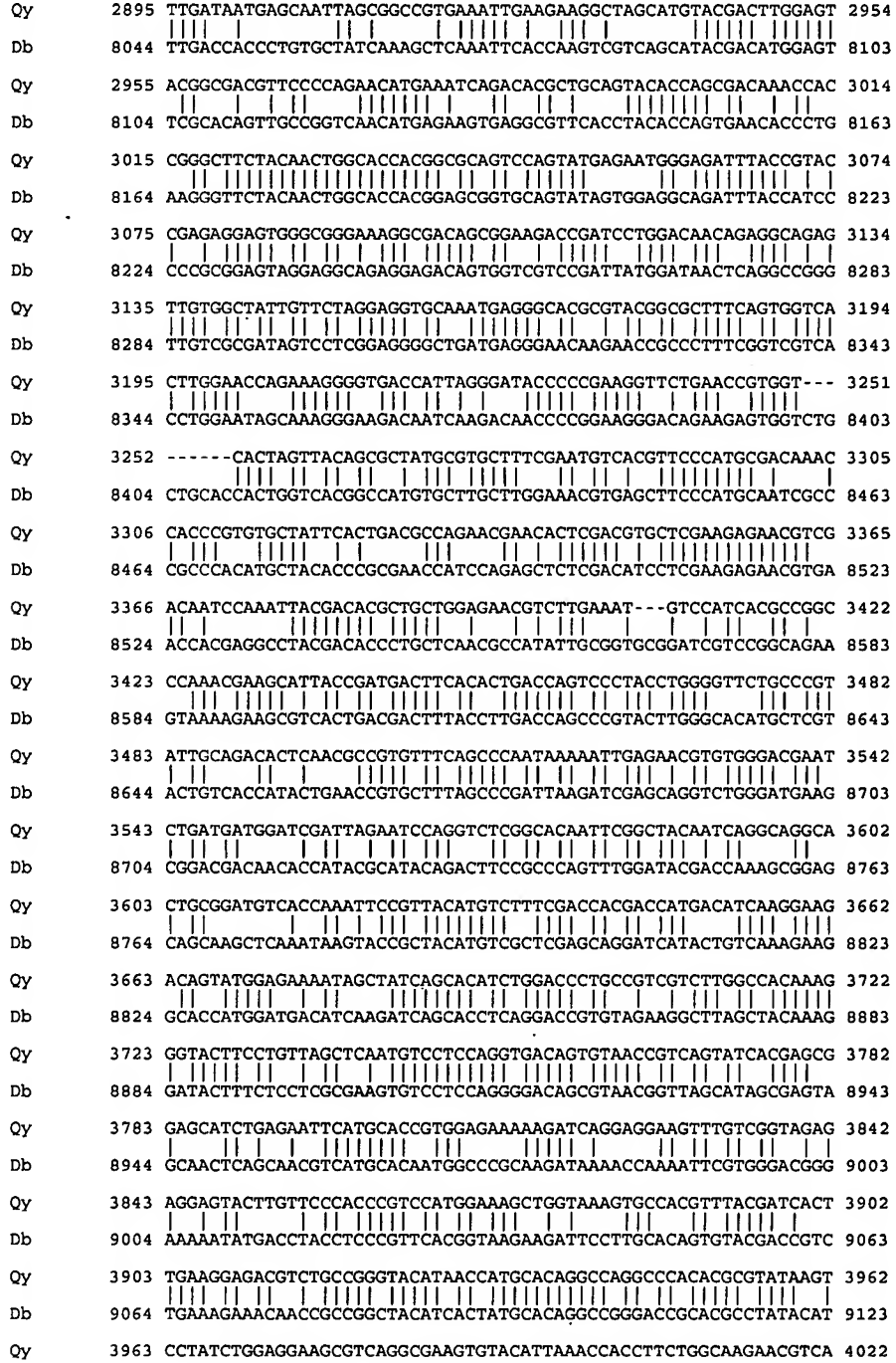
```
Qy       5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db      10984 ATGTATCCGACCGCGAAGGACAATGCCCTGTACATTCGCATTCGAGCACAGCAACCCTCC 11043

Qy       5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db      11044 AAGAGTCGACAGTTCATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTCAGCACCGCGA 11103

Qy       5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db      11104 GCCCACAGGCGAACTTCATTGTATCGCTGTGTGGTAAGAAGACAACATGCAATGCAGAAT 11163

Qy       6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db      11164 GCAAACCACCAGCTGATCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTCCAAG 11223

Qy       6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db      11224 CCGCCATCTCAAAAACTTCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11283

Qy       6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
Db      11284 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11343

Qy       6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
Db      11344 GACCGCTACGCCCCAATGACCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11403

Qy       6242 GCATAATGCCACGCGCCGCT----------------------------------TGACACT 6268
Db      11404 GCATAATGCATCAGGCTGGTATATTAGATCCCCGCTTACCGCGGGCAATATAGCAACACC 11463

Qy       6269 AAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGT 6318
Db      11464 AAAACTCGACGTATTTCCGAGGAAGCGCAGTGCATAATGCTGCGCAGTGT 11513
```

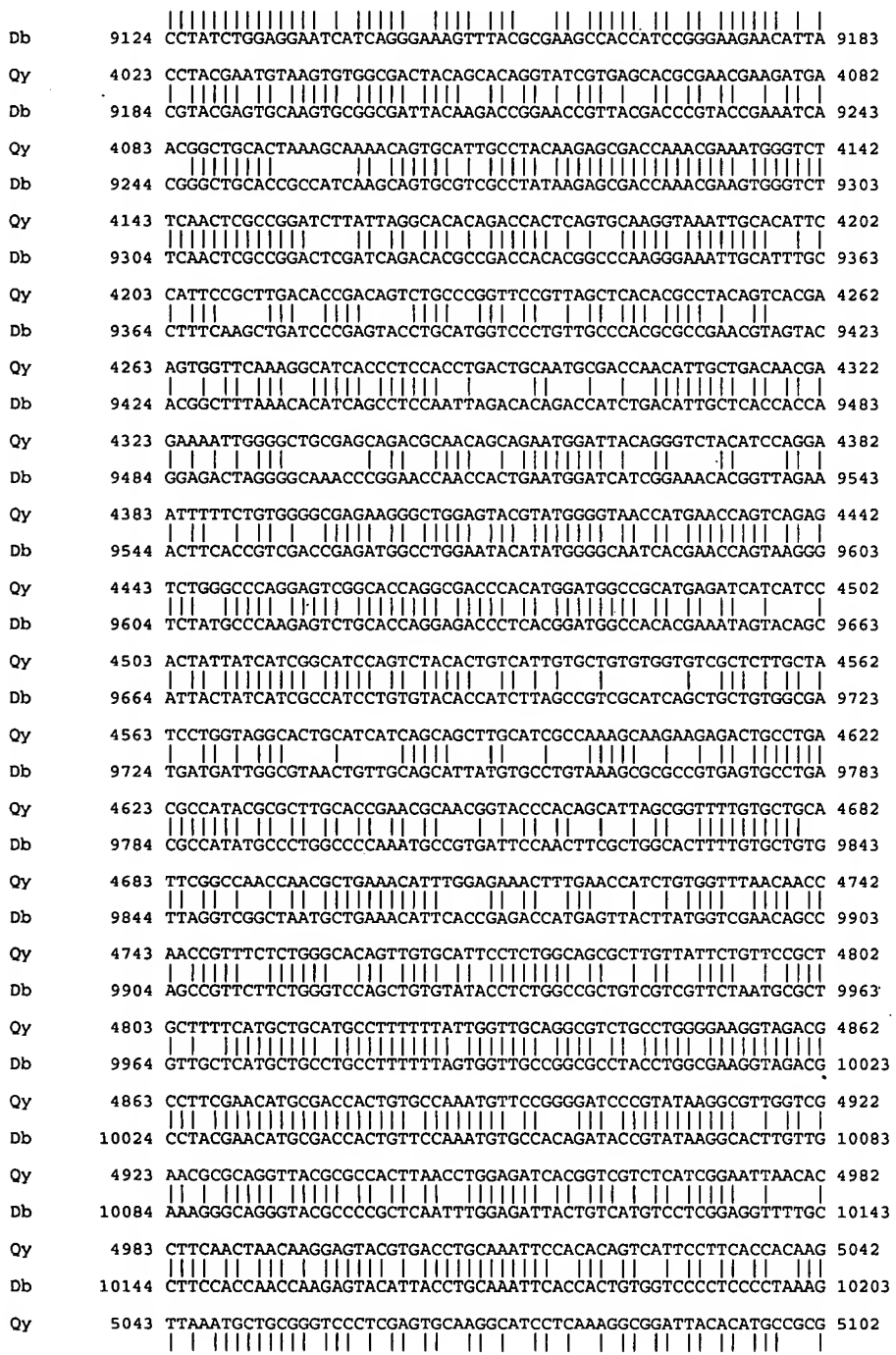
RESULT 6
AR096557
LOCUS AR096557 11663 bp DNA linear PAT 08-SEP-2000
DEFINITION Sequence 7 from patent US 6008035.
ACCESSION AR096557
VERSION AR096557.1 GI:10025467
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 11663)
AUTHORS Johnston,R.E., Davis,N.L. and Simpson,D.A.
TITLE System for the in vivo delivery and expression of heterologous genes in the bone marrow
JOURNAL Patent: US 6008035-A 7 28-DEC-1999;
FEATURES Location/Qualifiers
source 1. .11663
/organism="unknown"
BASE COUNT 3289 a 3105 c 2910 g 2359 t
ORIGIN

Query Match 37.0%; Score 2397.6; DB 6; Length 11663;
Best Local Similarity 64.6%; Pred. No. 0;
Matches 3752; Conservative 0; Mismatches 1984; Indels 74; Gaps 9;

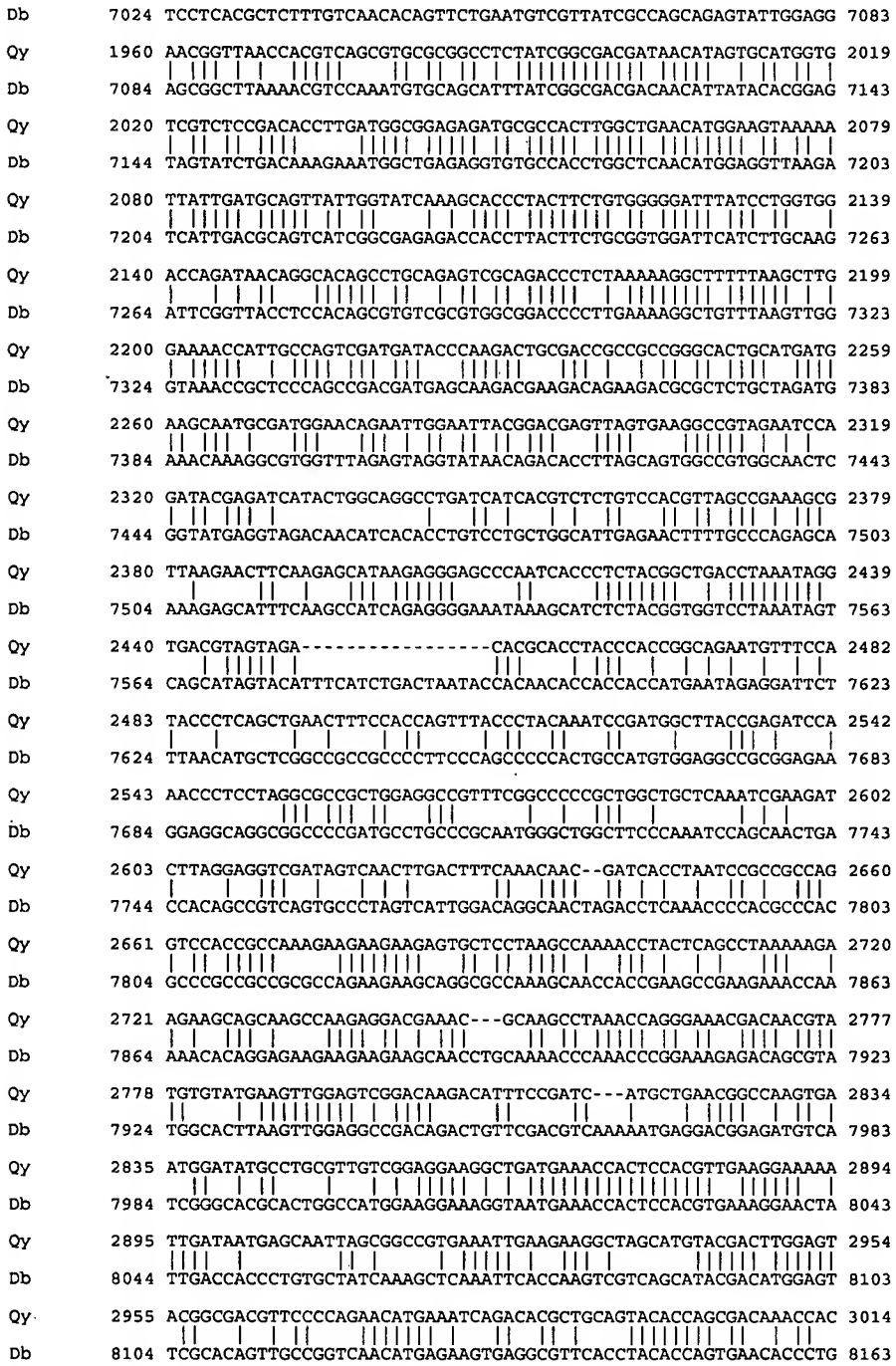
```
Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
Db       5707 ACTGTCTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5766

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
Db       5767 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTTCTGG 5826

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
Db       5827 AAAGAATCTACGCCCCGGTGCTCGACACGTCGAAAGAGGAACAGCTCAAACTCAGGTACC 5886

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
Db       5887 AGATGATGCCCACCGAAGCCAACAAAAGCAGGTACCAGTCTCGAAAAGTAGAAAACCAGA 5946

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
Db       5947 AAGCCATAACCACTGAGCGACTGCTTTCAGGGCTACGGCTGTATAACTCTGCCACAG--- 6003

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
Db       6004 ATCAGCCAGAATGCTATAAGATCACCTACCCGAAACCATCGTATTCCAGCAGTGTACCAG 6063

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
Db       6064 CGAACTACTCTGACCCAAAGTTTGCTGTAGCTGTTTGTAACAACTATCTGCATGAGAATT 6123

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
Db       6124 ACCCGACGGTAGCATCTTATCAGATCACCGACGAGTACGATGCTTACTTGGATATGGTAG 6183

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
Db       6184 ACGGGACAGTCGCTTGCCTAGATACTGCAACTTTTTGCCCCGCCAAGCTTAGAAGTTACC 6243

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
Db       6244 CGAAAAGACACGAGTATAGAGCCCCAAACATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6303

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
Db       6304 ACACGTTGCAAAACGTGCTCATTGCCGCGACTAAAAGAAACTGCAACGTCACACAAATGC 6363

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
Db       6364 GTGAACTGCCAACACTGGACTCAGCGACATTCAACGTTGAATGCTTTCGAAAATATGCAT 6423

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
Db       6424 GCAATGACGAGTATTGGGAGGAGTTTGCCCGAAAGCCAATTAGGATCACTACTGAGTTCG 6483

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
Db       6484 TTACCGCATACGTGGCCAGACTGAAAGGCCCTAAGGCCGCCGCACTGTTCGCAAAGACGC 6543

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
Db       6544 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6603

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
Db       6604 ACGTGAAAGTTACACCTGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6663

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
Db       6664 AAGCCGCAGAACCCCTGGCGACCGCTTACCTATGCGGGATCCACCGGGAGTTAGTGCGCA 6723

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
Db       6724 GGCTTACAGCCGTTTTGCTACCCAACATTCACACGCTCTTTGACATGTCGGCGGAGGACT 6783

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
Db       6784 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGTGACCCGGTACTGGAGACGGATATCG 6843

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
Db       6844 CCTCGTTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGCCTGATGATCTTGGAAG 6903

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
Db       6904 ACCTGGGTGTGGACCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 6963
```

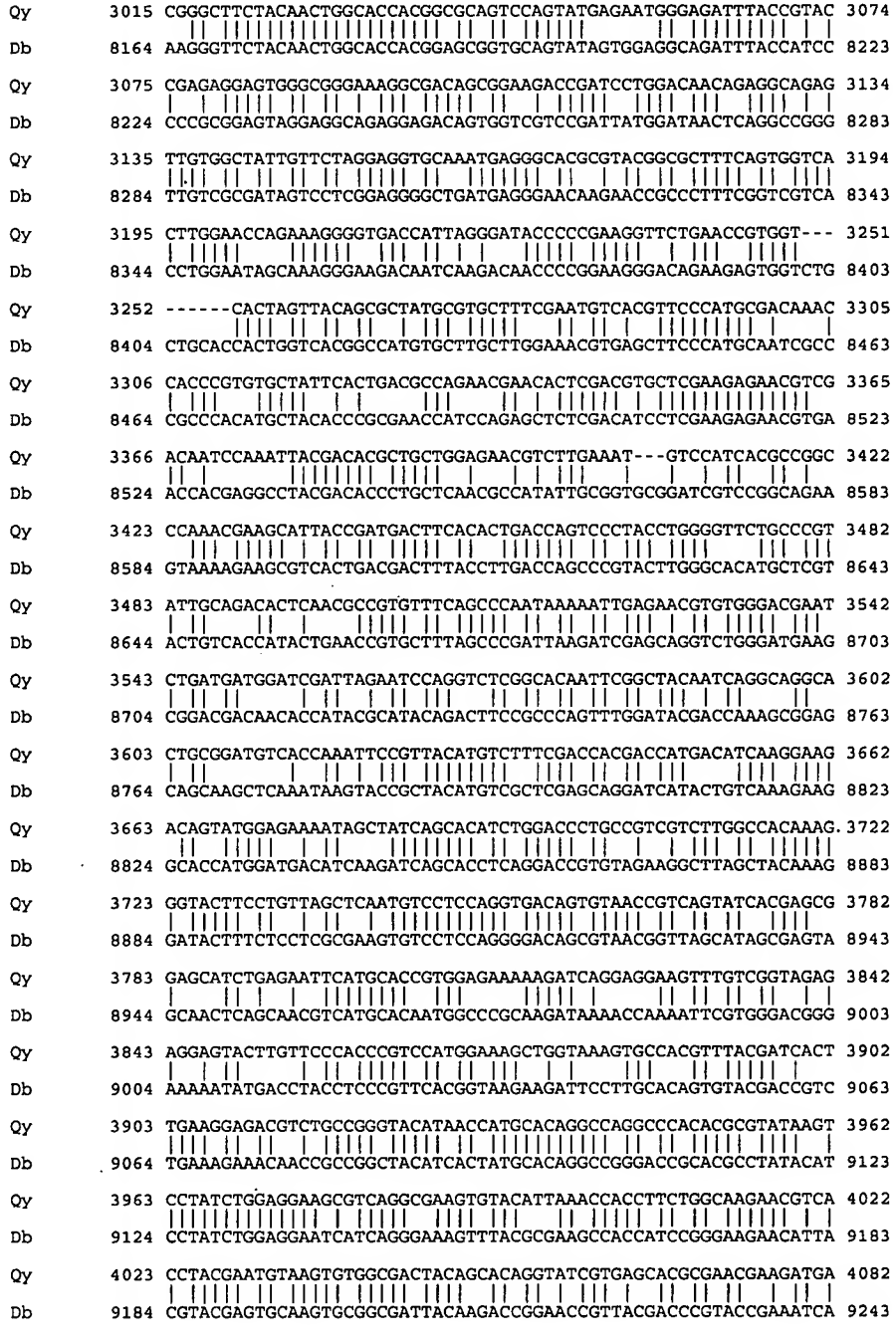


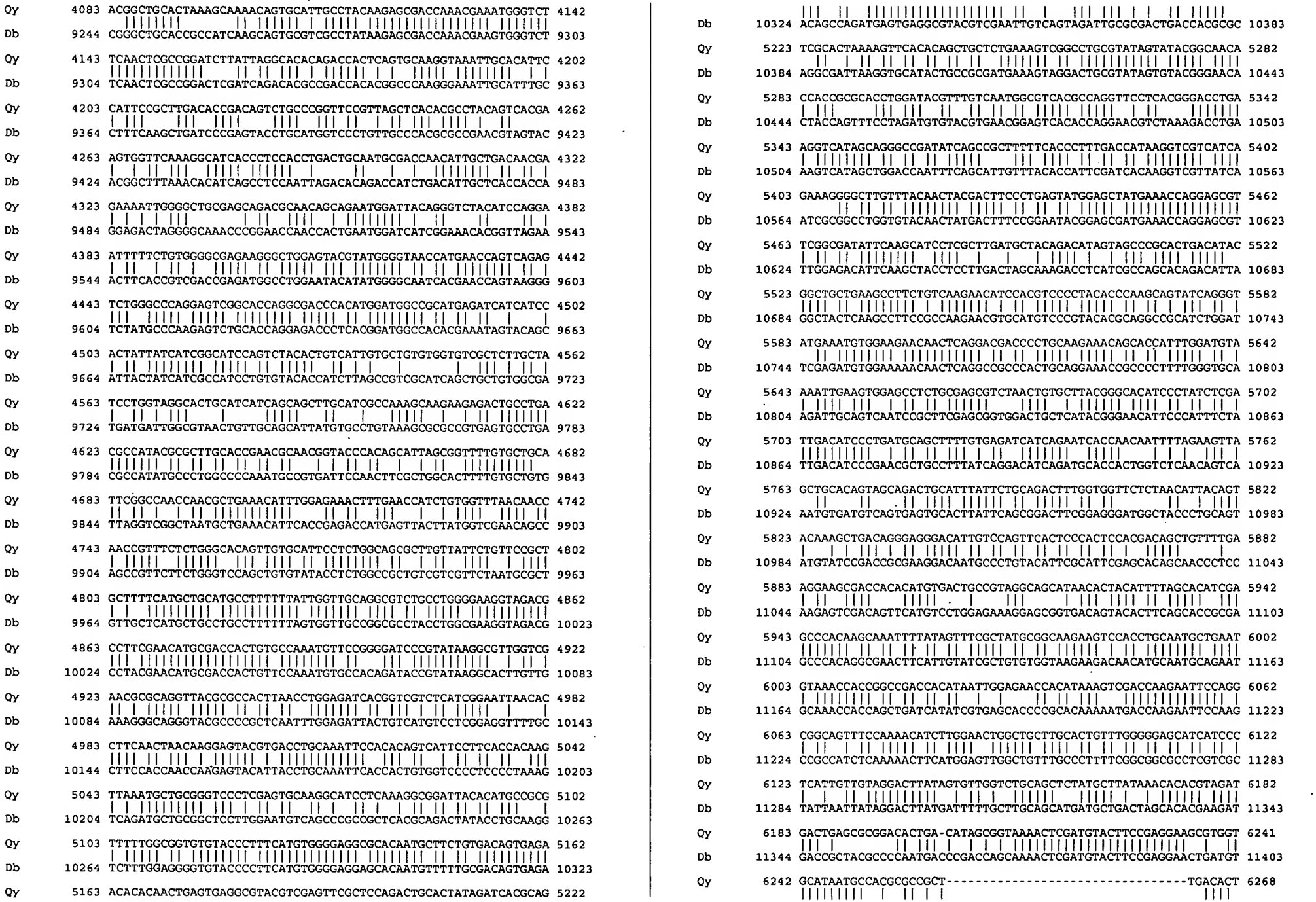



Db 9124 CCTATCTGGAGGAATCATCAGGGAAAGTTTACGCGAAGCCACCATCCGGGAAGAACATTA 9183
Qy 4023 CCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGATGA 4082
Db 9184 CGTACGAGTGCAAGTGCGGCGATTACAAGACCGGAACCGTTACGACCCGTACCGAAATCA 9243
Qy 4083 ACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGTCT 4142
Db 9244 CGGGCTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAACGAAGTGGGTCT 9303
Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9304 TCAACTCGCCGGACTCGATCAGACACGCCGACCACACGGCCCAAGGGAAATTGCATTTGC 9363
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9364 CTTTCAAGCTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAACGTAGTAC 9423
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9424 ACGGCTTTAAACACATCAGCCTCCAATTAGACACAGACCATCTGACATTGCTCACCACCA 9483
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9484 GGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCATCGGAAACACGGTTAGAA 9543
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9544 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGCAATCACGAACCAGTAAGGG 9603
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db 9604 TCTATGCCCAAGAGTCTGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9663
Qy 4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db 9664 ATTACTATCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTGCTGTGGCGA 9723
Qy 4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db 9724 TGATGATTGGCGTAACTGTTGCAGCATTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9783
Qy 4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db 9784 CGCCATATGCCCTGGCCCCAAATGCCGTGATTCCAACTTCGCTGGCACTTTTGTGCTGTG 9843
Qy 4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db 9844 TTAGGTCGGCTAATGCTGAAACATTCACCGAGACCATGAGTTACTTATGGTCGAACAGCC 9903
Qy 4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
Db 9904 AGCCGTTCTTCTGGGTCCAGCTGTGTATACCTCTGGCCGCTGTCGTCGTTCTAATGCGCT 9963
Qy 4803 GCTTTTCATGCTGCATGCCTTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
Db 9964 GTTGCTCATGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10023
Qy 4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
Db 10024 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10083
Qy 4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db 10084 AAAGGGCAGGGTACGCCCCGCTCAATTTGGAGATTACTGTCATGTCCTCGGAGGTTTTGC 10143
Qy 4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db 10144 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCTAAAG 10203
Qy 5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db 10204 TCAGATGCTGCGGCTCCTTGGAATGTCAGCCCGCCGCTCACGCAGACTATACCTGCAAGG 10263
Qy 5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db 10264 TCTTTGGAGGGGTGTACCCCTTCATGTGGGGAGGAGCACAATGTTTTTGCGACAGTGAGA 10323
Qy 5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db 10324 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGTAGATTGCGCGACTGACCACGCGC 10383
Qy 5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db 10384 AGGCGATTAAGGTGCATACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10443
Qy 5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db 10444 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACCTGA 10503
Qy 5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db 10504 AAGTCATAGCTGGACCAATTTCAGCATTGTTTACACCATTCGATCACAAGGTCGTTATCA 10563
Qy 5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db 10564 ATCGCGGCCTGGTGTACAACTATGACTTTCCGGAATACGGAGCGATGAAACCAGGAGCGT 10623
Qy 5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db 10624 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAAGACCTCATCGCCAGCACAGACATTA 10683
Qy 5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db 10684 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCTGGAT 10743
Qy 5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db 10744 TCGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCCCCTTTTGGGTGCA 10803
Qy 5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db 10804 AGATTGCAGTCAATCCGCTTCGAGCGGTGGACTGCTCATACGGGAACATTCCCATTTCTA 10863
Qy 5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db 10864 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10923
Qy 5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db 10924 AATGTGATGTCAGTGAGTGCACTTATTCAGCGGACTTCGGAGGGATGGCTACCCTGCAGT 10983
Qy 5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db 10984 ATGTATCCGACCGCGAAGGACAATGCCCTGTACATTCGCATTCGAGCACAGCAACCCTCC 11043
Qy 5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db 11044 AAGAGTCGACAGTTCATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTCAGCACCGCGA 11103
Qy 5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db 11104 GCCCACAGGCGAACTTCATTGTATCGCTGTGTGGTAAGAAGACAACATGCAATGCAGAAT 11163
Qy 6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db 11164 GCAAACCACCAGCTGATCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTCCAAG 11223
Qy 6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db 11224 CCGCCATCTCAAAAACTTCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11283
Qy 6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
Db 11284 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11343

```
Qy       6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
Db      11344 GACCGCTACGCCCCAATGACCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11403

Qy       6242 GCATAATGCCACGCGCCGCT----------------------------------TGACACT 6268
Db      11404 GCATAATGCATCAGGCTGGTATATTAGATCCCCGCTTACCGCGGGCAATATAGCAACACC 11463

Qy       6269 AAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGT 6318
Db      11464 AAAACTCGACGTATTTCCGAGGAAGCGCAGTGCATAATGCTGCGCAGTGT 11513


RESULT 7
BD071122
LOCUS       BD071122               11663 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  System for the in vivo delivery and expression of heterologous
            genes in the bone marrow.
ACCESSION   BD071122
VERSION     BD071122.1  GI:22616725
KEYWORDS    JP 2001515348-A/3.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 11663)
  AUTHORS   Johnston,R.E., Davis,N.L. and Simpson,D.A.
  TITLE     System for the in vivo delivery and expression of heterologous
            genes in the bone marrow
  JOURNAL   Patent: JP 2001515348-A 3 18-SEP-2001;
            UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
COMMENT     OS   Girdwood S.A.
            PN   JP 2001515348-A/3
            PD   18-SEP-2001
            PF   18-FEB-1998 JP 1998536760
            PR   19-FEB-1997 US    08/801263
            PI   ROBERT E JOHNSTON,NANCY L DAVIS,DENNIS A SIMPSON PC
            C12N15/86,C12N15/33,C12N7/01,C12N5/10,A61K39/12,A61K48/00 CC
            System for the in vivo delivery and expression of heterologous
            CC          genes in
            CC   the bone marrow
            FH   Key             Location/Qualifiers
            FT   source          1. .11663
            FT                   /organism='Girdwood S.A.'.
FEATURES             Location/Qualifiers
     source          1. .11663
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT     3289 a   3105 c   2910 g   2359 t
ORIGIN

  Query Match             37.0%;  Score 2397.6;  DB 6;  Length 11663;
  Best Local Similarity   64.6%;  Pred. No. 0;
  Matches 3752;  Conservative    0;  Mismatches 1984;  Indels   74;  Gaps    9;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
Db       5707 ACTGTCTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5766

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
Db       5767 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTTCTGG 5826

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
Db       5827 AAAGAATCTACGCCCCGGTGCTCGACACGTCGAAAGAGGAACAGCTCAAACTCAGGTACC 5886

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
Db       5887 AGATGATGCCCACCGAAGCCAACAAAAGCAGGTACCAGTCTCGAAAAGTAGAAAACCAGA 5946

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879

Db       5947 AAGCCATAACCACTGAGCGACTGCTTTCAGGGCTACGGCTGTATAACTCTGCCACAG--- 6003

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
Db       6004 ATCAGCCAGAATGCTATAAGATCACCTACCCGAAACCATCGTATTCCAGCAGTGTACCAG 6063

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
Db       6064 CGAACTACTCTGACCCAAAGTTTGCTGTAGCTGTTTGTAACAACTATCTGCATGAGAATT 6123

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
Db       6124 ACCCGACGGTAGCATCTTATCAGATCACCGACGAGTACGATGCTTACTTGGATATGGTAG 6183

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
Db       6184 ACGGGACAGTCGCTTGCCTAGATACTGCAACTTTTTGCCCCGCCAAGCTTAGAAGTTACC 6243

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
Db       6244 CGAAAAGACACGAGTATAGAGCCCCAAACATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6303

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
Db       6304 ACACGTTGCAAAACGTGCTCATTGCCGCGACTAAAAGAAACTGCAACGTCACACAAATGC 6363

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
Db       6364 GTGAACTGCCAACACTGGACTCAGCGACATTCAACGTTGAATGCTTTCGAAAATATGCAT 6423

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
Db       6424 GCAATGACGAGTATTGGGAGGAGTTTGCCCGAAAGCCAATTAGGATCACTACTGAGTTCG 6483

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
Db       6484 TTACCGCATACGTGGCCAGACTGAAAGGCCCTAAGGCCGCCGCACTGTTCGCAAAGACGC 6543

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
Db       6544 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6603

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
Db       6604 ACGTGAAAGTTACACCTGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6663

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
Db       6664 AAGCCGCAGAACCCCTGGCGACCGCTTACCTATGCGGGATCCACCGGGAGTTAGTGCGCA 6723

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
Db       6724 GGCTTACAGCCGTTTTGCTACCCAACATTCACACGCTCTTTGACATGTCGGCGGAGGACT 6783

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
Db       6784 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGTGACCCGGTACTGGAGACGGATATCG 6843

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
Db       6844 CCTCGTTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGCCTGATGATCTTGGAAG 6903

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
Db       6904 ACCTGGGTGTGGACCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 6963

Qy       1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
Db       6964 CCACCCATCTGCCCACGGGTACCCGTTTCAAATTCGGGGCGATGATGAAATCCGGAATGT 7023

Qy       1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
```

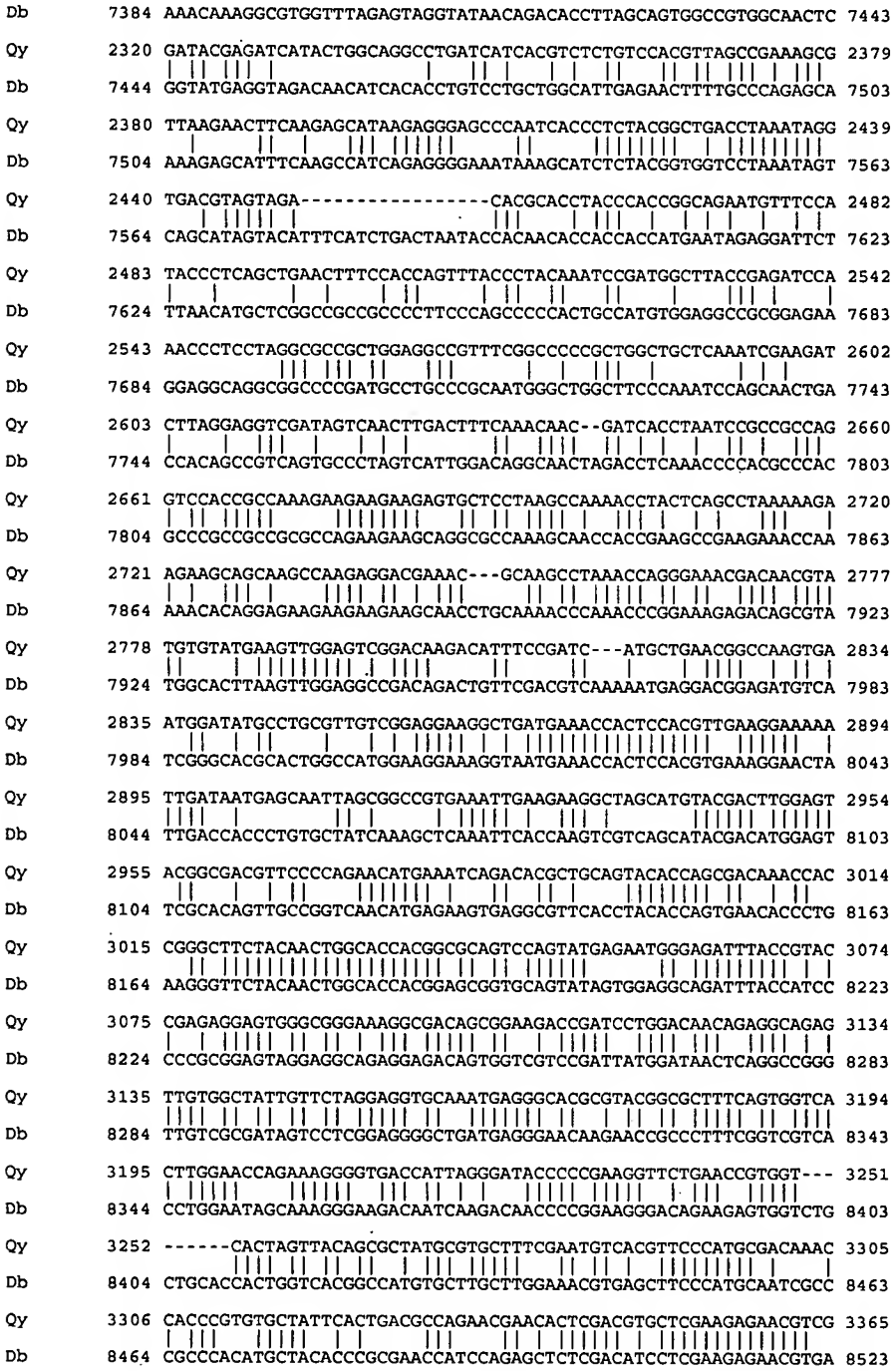


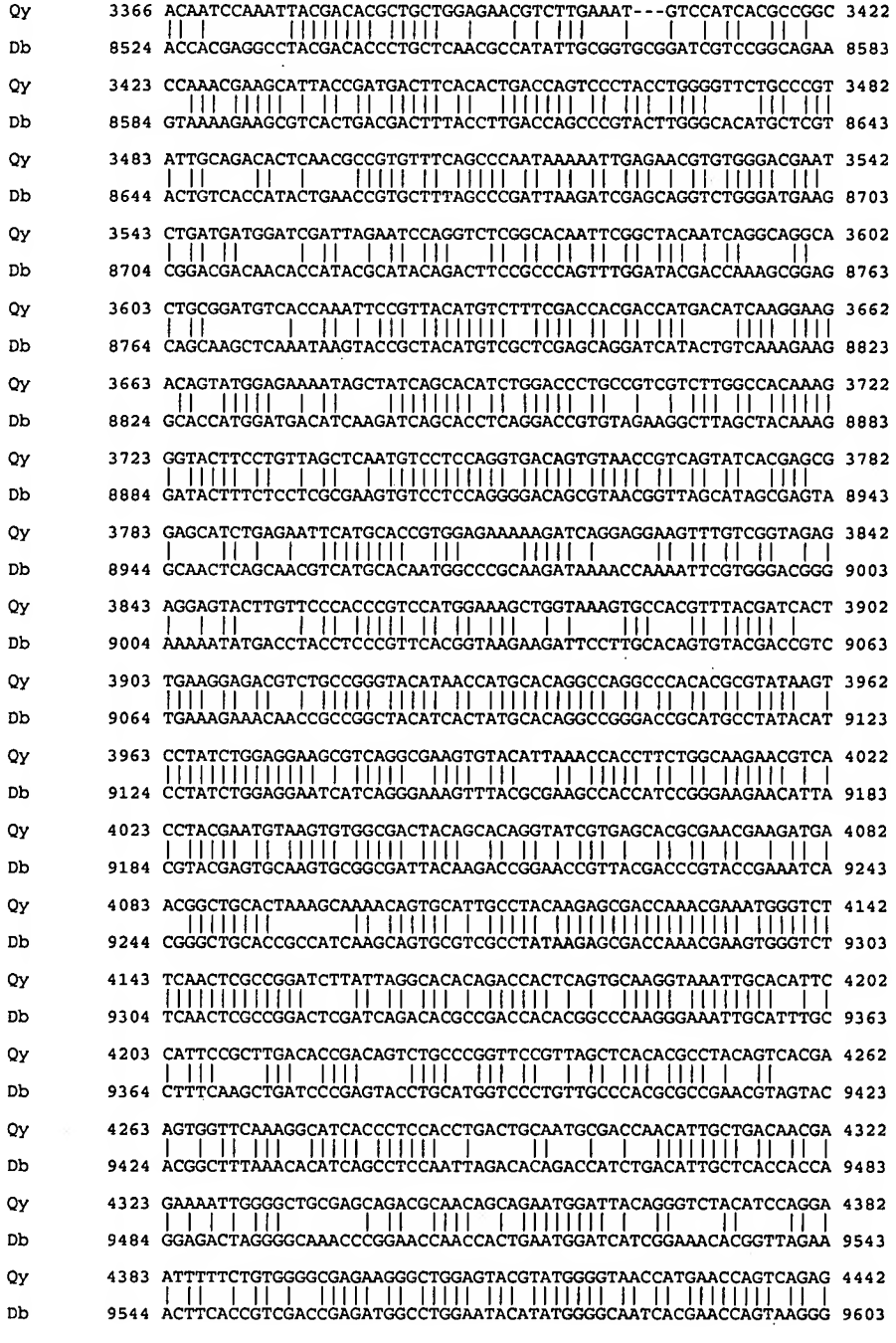


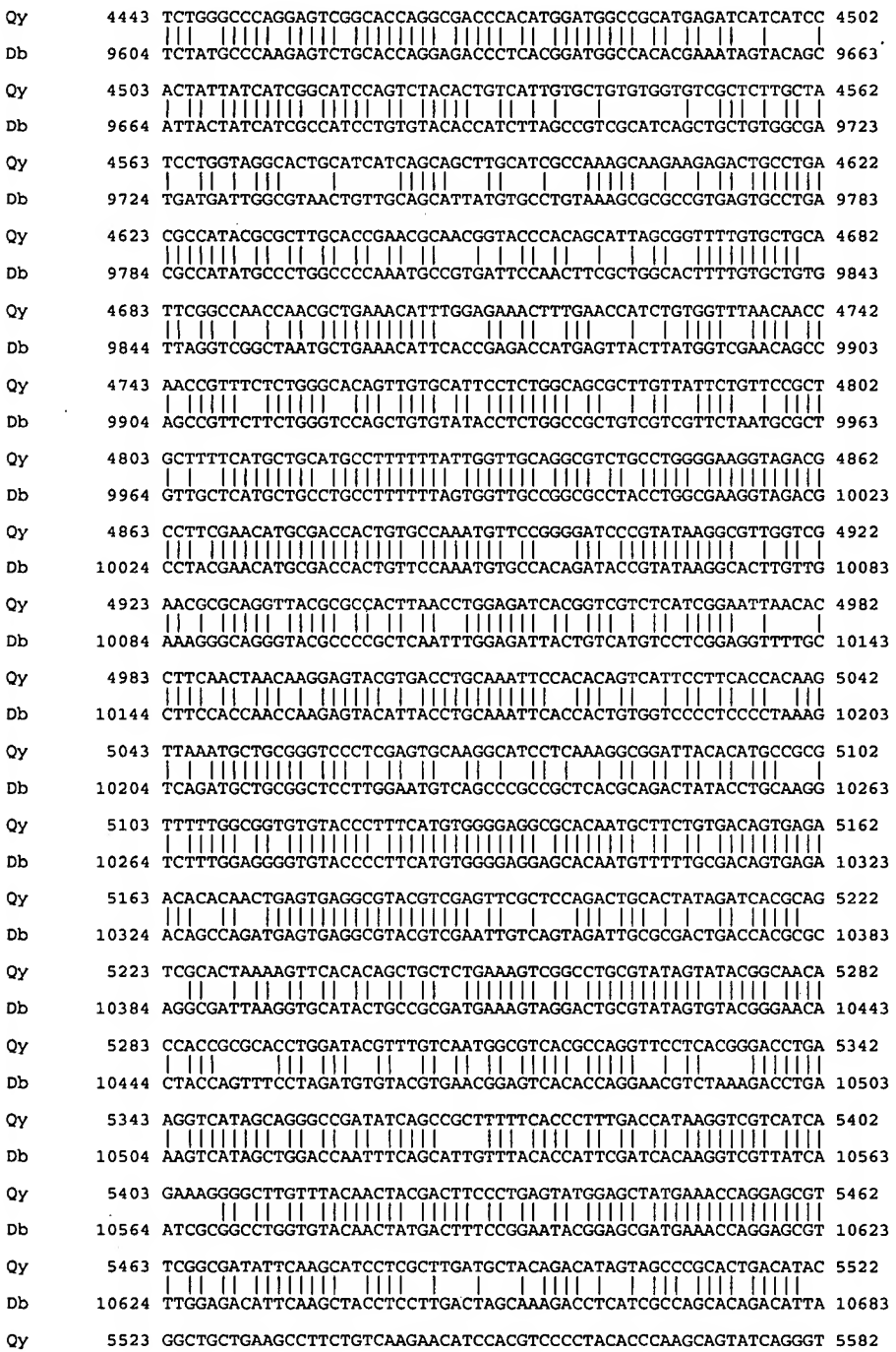


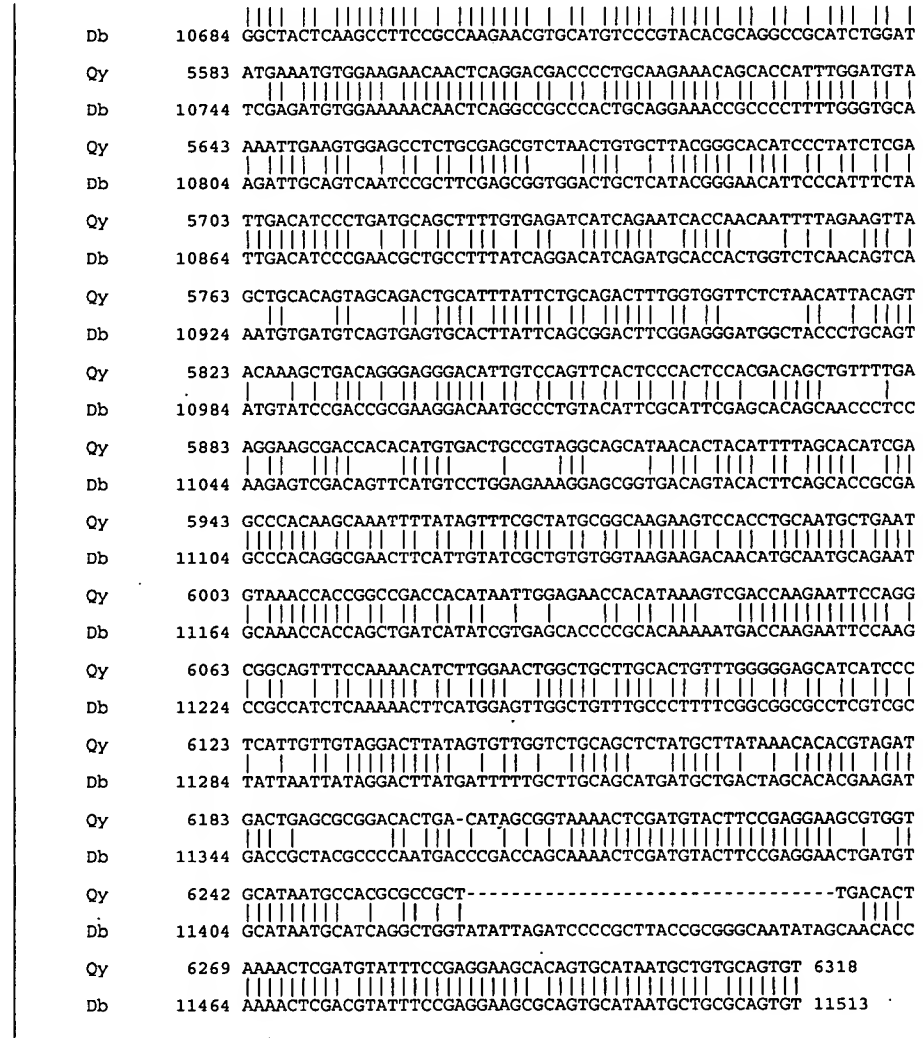
```
Db      11404 GCATAATGCATCAGGCTGGTATATTAGATCCCCGCTTACCGCGGGCAATATAGCAACACC 11463

Qy       6269 AAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGT 6318
              ||||||||| |||||||||||||||| |||||||||||||| |||||||
Db      11464 AAAACTCGACGTATTTCCGAGGAAGCGCAGTGCATAATGCTGCGCAGTGT 11513


RESULT 8
AR042409
LOCUS       AR042409               11663 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 1 from patent US 5811407.
ACCESSION   AR042409
VERSION     AR042409.1  GI:5962905
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 11663)
  AUTHORS   Johnston,R.E., Davis,N.L. and Simpson,D.A.
  TITLE     System for the in vivo delivery and expression of heterologous
            genes in the bone marrow
  JOURNAL   Patent: US 5811407-A 1 22-SEP-1998;
FEATURES             Location/Qualifiers
     source          1. .11663
                     /organism="unknown"
BASE COUNT     3291 a   3105 c   2907 g   2360 t
ORIGIN

  Query Match             36.9%;  Score 2396;  DB 6;  Length 11663;
  Best Local Similarity   64.6%;  Pred. No. 0;
  Matches 3751;  Conservative    0;  Mismatches 1985;  Indels   74;  Gaps    9;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
              |||| |     |  |  || | ||| || || ||  |||| |||||||  || ||| | |
Db       5707 ACTGTCTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5766

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
              || |||| || || |  ||      | ||   |||||| |  ||||| ||    || | |
Db       5767 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTTCTGG 5826

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
              | |     |||||||||   |||||        | |||| ||  | |  |     |
Db       5827 AAAGAATCTACGCCCCGGTGCTCGACACGTCGAAAGAGGAACAGCTCAAACTCAGGTACC 5886

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
              |  |    || | ||||  || | |||||||||| || || ||||||||||||||   ||
Db       5887 AGATGATGCCCACCGAAGCCAACAAAAGCAGGTACCAGTCTCGAAAAGTAGAAAACCAGA 5946

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
              |||| || ||  | |||||||  |||| ||  |  |    |||   ||  |  ||
Db       5947 AAGCCATAACCACTGAGCGACTGCTTTCAGGGCTACGACTGTATAACTCTGCCACAG--- 6003

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
              |||      || || || |   |||  || ||   ||||   || ||     |  |||
Db       6004 ATCAGCCAGAATGCTATAAGATCACCTACCCGAAACCATCGTATTCCAGCAGTGTACCAG 6063

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
                |  |   |    || || | || ||       || |||     | | || |||||||
Db       6064 CGAACTACTCTGACCCAAAGTTTGCTGTAGCTGTTTGTAACAACTATCTGCATGAGAATT 6123

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
              |||| || |||||   ||||  || || || || |||||||| ||  | || ||||| |
Db       6124 ACCCGACGGTAGCATCTTATCAGATCACCGACGAGTACGATGCTTACTTGGATATGGTAG 6183

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
              ||||  ||       || |||||||| || |||||||| || || || || ||||| ||||
Db       6184 ACGGGACAGTCGCTTGCCTAGATACTGCAACTTTTTGCCCCGCCAAGCTTAGAAGTTACC 6243

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
              | || |  ||    |||      ||| | ||  |    || || |||||  | || ||||
Db       6244 CGAAAAGACACGAGTATAGAGCCCCAAACATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6303

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
```

```
                | || || ||||| ||  |   ||| || |||||||| || |||||||| || |||||||
Db       6304 ACACGTTGCAAAACGTGCTCATTGCCGCGACTAAAAGAAACTGCAACGTCACACAAATGC 6363

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
              | ||| | ||      | || || ||| |||| || ||||| || ||     ||||| ||||
Db       6364 GTGAACTGCCAACACTGGACTCAGCGACATTCAACGTTGAATGCTTTCGAAAATATGCAT 6423

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
              ||||||| ||||| |||||     ||| |   || || ||| || | ||||| |||  ||
Db       6424 GCAATGACGAGTATTGGGAGGAGTTTGCCCGAAAGCCAATTAGGATCACTACTGAGTTCG 6483

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
              ||||   ||| ||| | |  |||||||| || || || || ||| |||| || || || |
Db       6484 TTACCGCATACGTGGCCAGACTGAAAGGCCCTAAGGCCGCCGCACTGTTCGCAAAGACGC 6543

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
              ||||| |     || |||| ||  | || ||||||  |||||||||||   | || ||||
Db       6544 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6603

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
              | || |||||||| || |||||||||| ||||| ||  | || || || || || || | |
Db       6604 ACGTGAAAGTTACACCTGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6663

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
              | || ||||| ||||| || |||||||||| |||||||||| ||||| ||||| || |
Db       6664 AAGCCGCAGAACCCCTGGCGACCGCTTACCTATGCGGGATCCACCGGGAGTTAGTGCGCA 6723

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
              | || |  || ||  | || || || || ||||| |||||||| ||||| || | |
Db       6724 GGCTTACAGCCGTTTTGCTACCCAACATTCACACGCTCTTTGACATGTCGGCGGAGGACT 6783

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
              ||||||| || || || ||||| ||| | || || ||||| ||| |||| ||||| ||||
Db       6784 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGTGACCCGGTACTGGAGACGGATATCG 6843

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
              | ||||| || |||||| |||||||||||| ||  |  | |  |||||||| | || |
Db       6844 CCTCGTTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGCCTGATGATCTTGGAAG 6903

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
              || | ||||| |||||||| || | || ||||| |||   || || ||  | ||  |||
Db       6904 ACCTGGGTGTGGACCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 6963

Qy       1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
              |     || || || || || ||  | || ||||| || || ||||||||||||||  ||||
Db       6964 CCACCCATCTGCCCACGGGTACCCGTTTCAAATTCGGGGCGATGATGAAATCCGGAATGT 7023

Qy       1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
              || | ||||| |||||||||||  |  ||| || | || || |||||||| |   |
Db       7024 TCCTCACGCTCTTTGTCAACACAGTTCTGAATGTCGTTATCGCCAGCAGAGTATTGGAGG 7083

Qy       1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
              | ||| | |  |||||     || || || | |||||||||| ||||  | || || |
Db       7084 AGCGGCTTAAAACGTCCAAATGTGCAGCATTTATCGGCGACGACAACATTATACACGGAG 7143

Qy       2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
              | || || ||||      ||||| ||||| || ||||| ||||| |||||||| || || |
Db       7144 TAGTATCTGACAAAGAAATGGCTGAGAGGTGTGCCACCTGGCTCAACATGGAGGTTAAGA 7203

Qy       2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139
              | ||||| ||||| || ||    | | |||| |||||||| || ||||| ||| ||   |
Db       7204 TCATTGACGCAGTCATCGGCGAGAGACCACCTTACTTCTGCGGTGGATTCATCTTGCAAG 7263

Qy       2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199
              |   | | ||   |||||| || | || || ||||| | |||||||| |||||| | |
Db       7264 ATTCGGTTACCTCCACAGCGTGTCGCGTGGCGGACCCCTTGAAAAGGCTGTTTAAGTTGG 7323

Qy       2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
              | |||||  | |||| ||| |||   ||||||   ||| |  | || || ||||  ||||
Db       7324 GTAAACCGCTCCCAGCCGACGATGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 7383

Qy       2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319
              || ||| |   |||   ||| | || || || |||   ||||    |||||| | | |
```



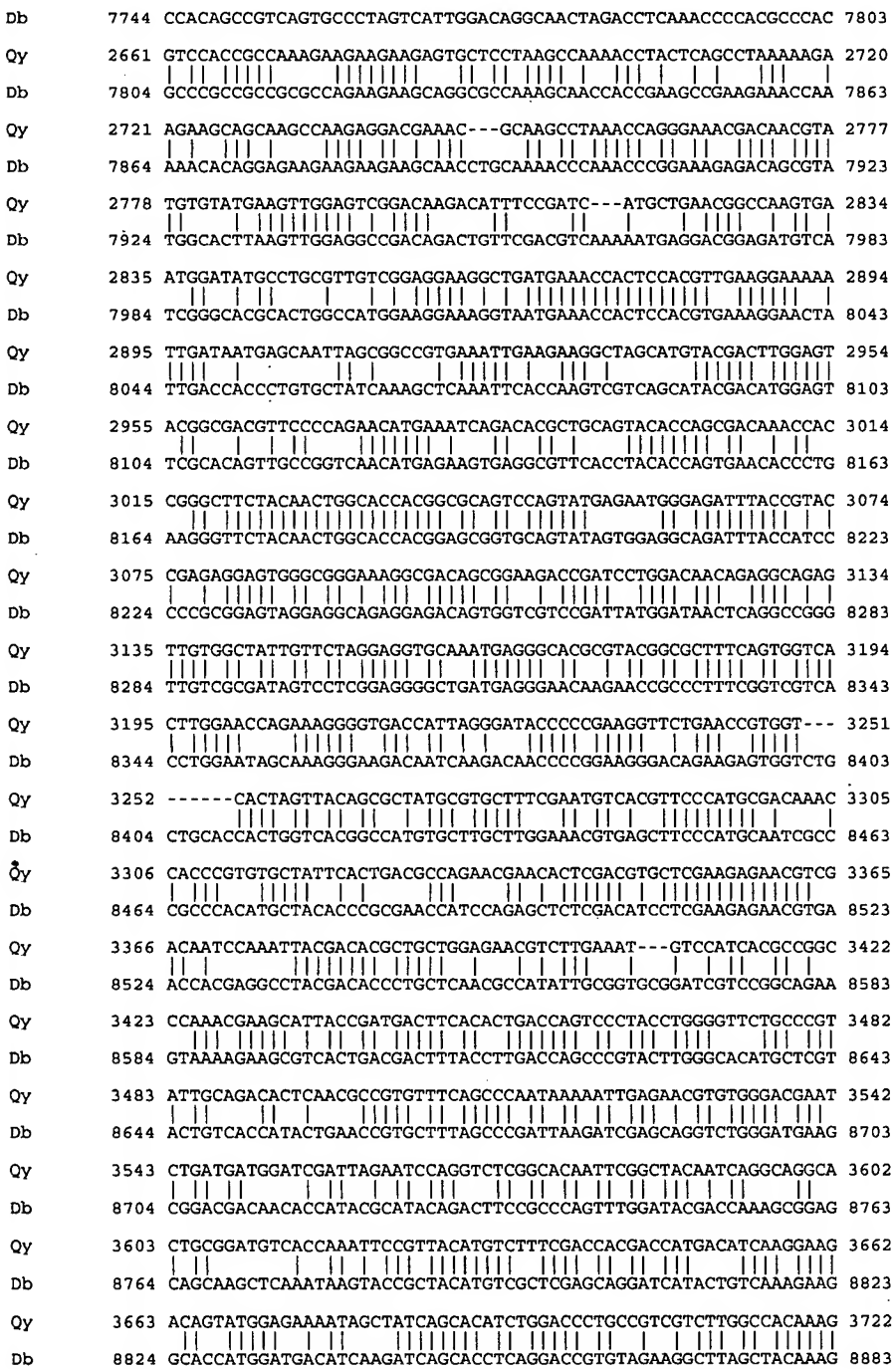






```
RESULT 9
AR096555
LOCUS       AR096555               11663 bp    DNA     linear   PAT 08-SEP-2000
DEFINITION  Sequence 1 from patent US 6008035.
ACCESSION   AR096555
VERSION     AR096555.1  GI:10025463
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 11663)
  AUTHORS   Johnston,R.E., Davis,N.L. and Simpson,D.A.
  TITLE     System for the in vivo delivery and expression of heterologous
            genes in the bone marrow
  JOURNAL   Patent: US 6008035-A 1 28-DEC-1999;
FEATURES             Location/Qualifiers
     source          1. .11663
```

```
                     /organism="unknown"
BASE COUNT     3291 a   3105 c   2907 g   2360 t
ORIGIN

  Query Match           36.9%;  Score 2396;  DB 6;  Length 11663;
  Best Local Similarity  64.6%;  Pred. No. 0;
  Matches 3751;  Conservative    0;  Mismatches 1985;  Indels   74;  Gaps    9;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
              |||| |    |  |  || | ||| || || ||  |||| |||||||  || ||| | |
Db       5707 ACTGTCTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5766

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
              || |||| || || |  ||      | ||   ||||||  |  ||||| ||    || | |
Db       5767 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTTCTGG 5826

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
              | |     |||||||||   |||||        | |||| ||  | |  |      |
Db       5827 AAAGAATCTACGCCCCGGTGCTCGACACGTCGAAAGAGGAACAGCTCAAACTCAGGTACC 5886

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
              |  |     || | ||||  || | |||||||||| || || ||||||||||||||   ||
Db       5887 AGATGATGCCCACCGAAGCCAACAAAAGCAGGTACCAGTCTCGAAAAGTAGAAAACCAGA 5946

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
              |||| || ||  | |||||||||  |||| ||  |  |     |||    ||  |   ||
Db       5947 AAGCCATAACCACTGAGCGACTGCTTTCAGGGCTACGACTGTATAACTCTGCCACAG--- 6003

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
              |||      || || || |   |||  || ||    ||||   || ||     |  |||
Db       6004 ATCAGCCAGAATGCTATAAGATCACCTACCCGAAACCATCGTATTCCAGCAGTGTACCAG 6063

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
                |  |   |     || || | || ||      || |||     | | || |||||||
Db       6064 CGAACTACTCTGACCCAAAGTTTGCTGTAGCTGTTTGTAACAACTATCTGCATGAGAATT 6123

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
              |||| || |||||   ||||    || || || || |||||||| ||  | || ||||| |
Db       6124 ACCCGACGGTAGCATCTTATCAGATCACCGACGAGTACGATGCTTACTTGGATATGGTAG 6183

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
              ||||  ||      || |||||||| || |||||||| || || || || ||||| ||||
Db       6184 ACGGGACAGTCGCTTGCCTAGATACTGCAACTTTTTGCCCCGCCAAGCTTAGAAGTTACC 6243

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
              | || |  ||     |||       ||| | ||  |    || || |||||  | || ||||
Db       6244 CGAAAAGACACGAGTATAGAGCCCCAAACATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6303

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
              | || || ||||| ||   |   ||| || |||||||| || |||||||| || |||||||
Db       6304 ACACGTTGCAAAACGTGCTCATTGCCGCGACTAAAAGAAACTGCAACGTCACACAAATGC 6363

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
              | ||| | ||     | || || ||| |||| || ||||| || ||     ||||| ||||
Db       6364 GTGAACTGCCAACACTGGACTCAGCGACATTCAACGTTGAATGCTTTCGAAAATATGCAT 6423

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
              |||||||  |||||  |||||    |||   |   || || ||| || | |||||  |||   ||
Db       6424 GCAATGACGAGTATTGGGAGGAGTTTGCCCGAAAGCCAATTAGGATCACTACTGAGTTCG 6483

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
              ||||   ||| ||| | |  |||||||| || || || ||| |||| || || || | |
Db       6484 TTACCGCATACGTGGCCAGACTGAAAGGCCCTAAGGCCGCCGCACTGTTCGCAAAGACGC 6543

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
              |||||  |    || ||||| ||  | || ||||||  ||||||||||||   | || ||||
Db       6544 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6603

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
              | || |||||||| || |||||||||||| |||||  ||  | || || || || || || | |
Db       6604 ACGTGAAAGTTACACCTGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6663

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
              | || |||||  ||||| || |||||||||| |||||||||| |||||  ||||| || | |
Db       6664 AAGCCGCAGAACCCCTGGCGACCGCTTACCTATGCGGGATCCACCGGGAGTTAGTGCGCA 6723

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
              | || |  || ||  | || || || || || || |||||  |||||||| ||||  || |
Db       6724 GGCTTACAGCCGTTTTGCTACCCAACATTCACACGCTCTTTGACATGTCGGCGGAGGACT 6783

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
              ||||||| || || || |||||  ||| | || || |||||  |||  ||||  |||||  ||||
Db       6784 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGTGACCCGGTACTGGAGACGGATATCG 6843

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
              | ||||| || |||||| ||||||||||||| ||  |  | |  ||||||| | || |
Db       6844 CCTCGTTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGCCTGATGATCTTGGAAG 6903

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
              || | ||||| |||||||| ||  | || |||||  |||    || || ||  | ||  |||
Db       6904 ACCTGGGTGTGGACCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 6963

Qy       1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
              |    || || || || || ||  | || |||||  || || ||||||||||||||  ||||
Db       6964 CCACCCATCTGCCCACGGGTACCCGTTTCAAATTCGGGGCGATGATGAAATCCGGAATGT 7023

Qy       1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
              || | ||||| |||||||||||| |  | ||| || | || || ||||||||| |    |
Db       7024 TCCTCACGCTCTTTGTCAACACAGTTCTGAATGTCGTTATCGCCAGCAGAGTATTGGAGG 7083

Qy       1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
              | ||| | |||||    || || | |||||||||||| ||||| | || || | || || |
Db       7084 AGCGGCTTAAAACGTCCAAATGTGCAGCATTTATCGGCGACGACAACATTATACACGGAG 7143

Qy       2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
              | || || ||||      ||||| ||||| || ||||| ||||| |||||||| || || |
Db       7144 TAGTATCTGACAAAGAAATGGCTGAGAGGTGTGCCACCTGGCTCAACATGGAGGTTAAGA 7203

Qy       2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139
              | ||||| ||||| || ||    | | |||| |||||||| || |||||  ||| ||   |
Db       7204 TCATTGACGCAGTCATCGGCGAGAGACCACCTTACTTCTGCGGTGGATTCATCTTGCAAG 7263

Qy       2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199
              |   | | ||   |||||| || | || || ||||  | |||||||| |||||| | |
Db       7264 ATTCGGTTACCTCCACAGCGTGTCGCGTGGCGGACCCCTTGAAAAGGCTGTTTAAGTTGG 7323

Qy       2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
              | |||||  | |||| ||| |||    ||||||   ||| |  | || || ||||  ||||
Db       7324 GTAAACCGCTCCCAGCCGACGATGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 7383

Qy       2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319
              || ||| |   |||   ||| | || || || |||   ||||    ||||||  | | |
Db       7384 AAACAAAGGCGTGGTTTAGAGTAGGTATAACAGACACCTTAGCAGTGGCCGTGGCAACTC 7443

Qy       2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379
              | || ||| |              |  || |  | |  ||   || ||| | |||  | |||
Db       7444 GGTATGAGGTAGACAACATCACACCTGTCCTGCTGGCATTGAGAACTTTTGCCCAGAGCA 7503

Qy       2380 TTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGG 2439
                |   ||  |  |||  ||||||    ||      ||||||||  | |||||||||
Db       7504 AAAGAGCATTTCAAGCCATCAGAGGGGAAATAAAGCATCTCTACGGTGGTCCTAAATAGT 7563

Qy       2440 TGACGTAGTAGA----------------CACGCACCTACCCACCGGCAGAATGTTTCCA 2482
                | |||||  |                |||    | |||  |  | | |  | | |
Db       7564 CAGCATAGTACATTTCATCTGACTAATACCACAACACCACCACCATGAATAGAGGATTCT 7623

Qy       2483 TACCCTCAGCTGAACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCA 2542
              |  |      |  |  |  |||  |  || ||  ||    |  ||| |   |
Db       7624 TTAACATGCTCGGCCGCCGCCCCTTCCCAGCCCCCACTGCCATGTGGAGGCCGCGGAGAA 7683

Qy       2543 AACCCTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGAT 2602
                  ||| ||| ||  |||    | |  |||  |    | | |
Db       7684 GGAGGCAGGCGGCCCCGATGCCTGCCCGCAATGGGCTGGCTTCCCAAATCCAGCAACTGA 7743

Qy       2603 CTTAGGAGGTCGATAGTCAACTTGACTTTCAAACAAC--GATCACCTAATCCGCCGCCAG 2660
              |    |  |||  |  | | |       || ||||  || | |  |  || |  |||
```

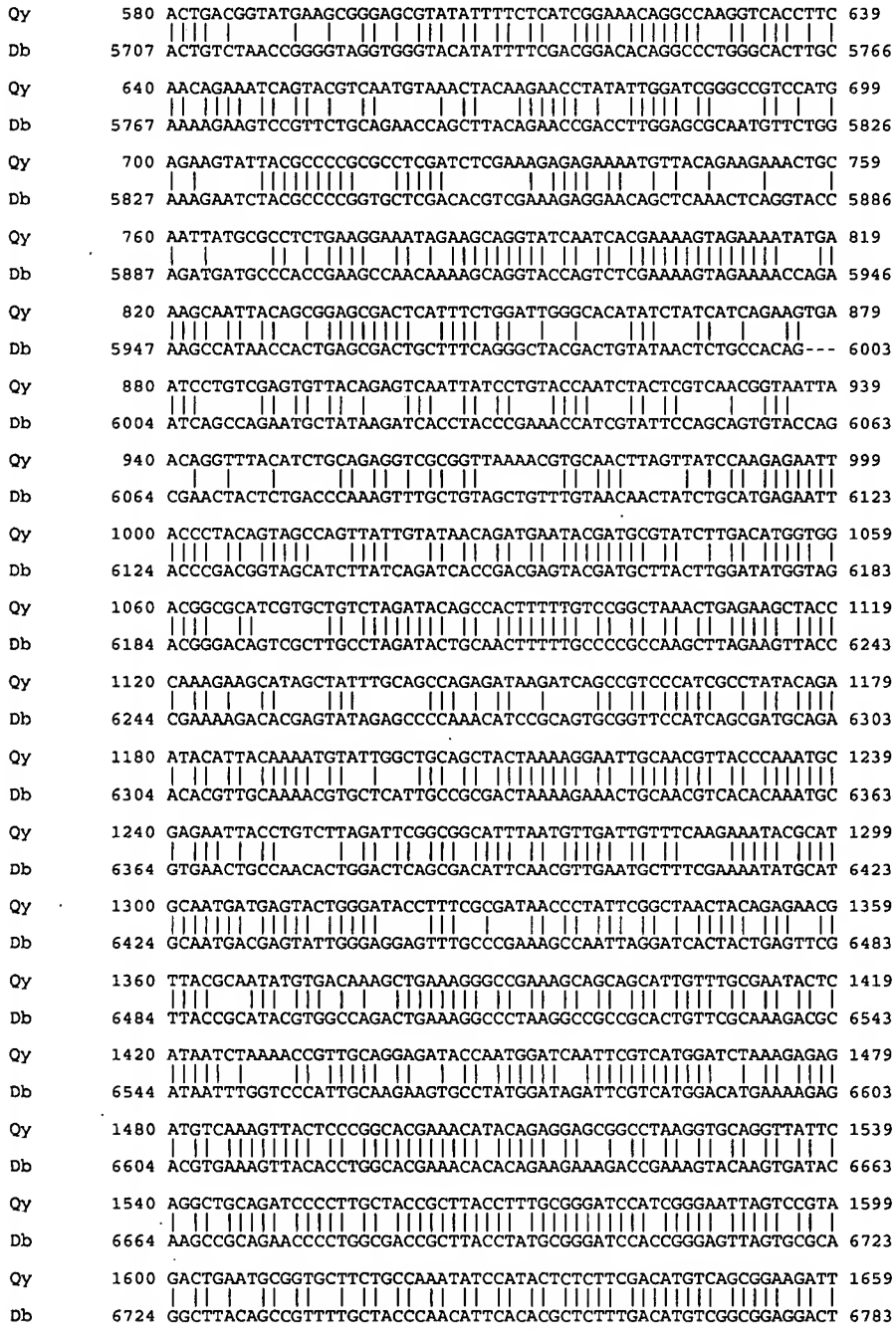

Db 7744 CCACAGCCGTCAGTGCCCTAGTCATTGGACAGGCAACTAGACCTCAAACCCCACGCCCAC 7803
Qy 2661 GTCCACCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGA 2720
Db 7804 GCCCGCCGCCGCGCCAGAAGAAGCAGGCGCCAAAGCAACCACCGAAGCCGAAGAAACCAA 7863
Qy 2721 AGAAGCAGCAAGCCAAGAGGACGAAAC---GCAAGCCTAAACCAGGGAAACGACAACGTA 2777
Db 7864 AAACACAGGAGAAGAAGAAGAAGCAACCTGCAAAACCCAAACCCGGAAAGAGACAGCGTA 7923
Qy 2778 TGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATC---ATGCTGAACGGCCAAGTGA 2834
Db 7924 TGGCACTTAAGTTGGAGGCCGACAGACTGTTCGACGTCAAAAATGAGGACGGAGATGTCA 7983
Qy 2835 ATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACTCCACGTTGAAGGAAAAA 2894
Db 7984 TCGGGCACGCACTGGCCATGGAAGGAAAGGTAATGAAACCACTCCACGTGAAAGGAACTA 8043
Qy 2895 TTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAGCATGTACGACTTGGAGT 2954
Db 8044 TTGACCACCCTGTGCTATCAAAGCTCAAATTCACCAAGTCGTCAGCATACGACATGGAGT 8103
Qy 2955 ACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTACACCAGCGACAAACCAC 3014
Db 8104 TCGCACAGTTGCCGGTCAACATGAGAAGTGAGGCGTTCACCTACACCAGTGAACACCCTG 8163
Qy 3015 CGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAATGGGAGATTTACCGTAC 3074
Db 8164 AAGGGTTCTACAACTGGCACCACGGAGCGGTGCAGTATAGTGGAGGCAGATTTACCATCC 8223
Qy 3075 CGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGACAACAGAGGCAGAG 3134
Db 8224 CCCGCGGAGTAGGAGGCAGAGGAGACAGTGGTCGTCCGATTATGGATAACTCAGGCCGGG 8283
Qy 3135 TTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGCGCTTTCAGTGGTCA 3194
Db 8284 TTGTCGCGATAGTCCTCGGAGGGGCTGATGAGGGAACAAGAACCGCCCTTTCGGTCGTCA 8343
Qy 3195 CTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTCTGAACCGTGGT--- 3251
Db 8344 CCTGGAATAGCAAAGGGAAGACAATCAAGACAACCCCGGAAGGGACAGAAGAGTGGTCTG 8403
Qy 3252 ------CACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGACAAAC 3305
Db 8404 CTGCACCACTGGTCACGGCCATGTGCTTGCTTGGAAACGTGAGCTTCCCATGCAATCGCC 8463
Qy 3306 CACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAACGTCG 3365
Db 8464 CGCCCACATGCTACACCCGCGAACCATCCAGAGCTCTCGACATCCTCGAAGAGAACGTGA 8523
Qy 3366 ACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAAT---GTCCATCACGCCGGC 3422
Db 8524 ACCACGAGGCCTACGACACCCTGCTCAACGCCATATTGCGGTGCGGATCGTCCGGCAGAA 8583
Qy 3423 CCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCCCGT 3482
Db 8584 GTAAAAGAAGCGTCACTGACGACTTTACCTTGACCAGCCCGTACTTGGGCACATGCTCGT 8643
Qy 3483 ATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACGAAT 3542
Db 8644 ACTGTCACCATACTGAACCGTGCTTTAGCCCGATTAAGATCGAGCAGGTCTGGGATGAAG 8703
Qy 3543 CTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAGGCA 3602
Db 8704 CGGACGACAACACCATACGCATACAGACTTCCGCCCAGTTTGGATACGACCAAAGCGGAG 8763
Qy 3603 CTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAAGGAAG 3662
Db 8764 CAGCAAGCTCAAATAAGTACCGCTACATGTCGCTCGAGCAGGATCATACTGTCAAAGAAG 8823
Qy 3663 ACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCACAAAG 3722
Db 8824 GCACCATGGATGACATCAAGATCAGCACCTCAGGACCGTGTAGAAGGCTTAGCTACAAAG 8883
Qy 3723 GGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCACGAGCG 3782
Db 8884 GATACTTTCTCCTCGCGAAGTGTCCTCCAGGGGACAGCGTAACGGTTAGCATAGCGAGTA 8943
Qy 3783 GAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTAGAG 3842
Db 8944 GCAACTCAGCAACGTCATGCACAATGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGG 9003
Qy 3843 AGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATCACT 3902
Db 9004 AAAAATATGACCTACCTCCCGTTCACGGTAAGAAGATTCCTTGCACAGTGTACGACCGTC 9063
Qy 3903 TGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAAGT 3962
Db 9064 TGAAAGAAACAACCGCCGGCTACATCACTATGCACAGGCCGGGACCGCATGCCTATACAT 9123
Qy 3963 CCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGTCA 4022
Db 9124 CCTATCTGGAGGAATCATCAGGGAAAGTTTACGCGAAGCCACCATCCGGGAAGAACATTA 9183
Qy 4023 CCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGATGA 4082
Db 9184 CGTACGAGTGCAAGTGCGGCGATTACAAGACCGGAACCGTTACGACCCGTACCGAAATCA 9243
Qy 4083 ACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGTCT 4142
Db 9244 CGGGCTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAACGAAGTGGGTCT 9303
Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9304 TCAACTCGCCGGACTCGATCAGACACGCCGACCACACGGCCCAAGGGAAATTGCATTTGC 9363
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9364 CTTTCAAGCTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAACGTAGTAC 9423
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9424 ACGGCTTTAAACACATCAGCCTCCAATTAGACACAGACCATCTGACATTGCTCACCACCA 9483
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9484 GGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCATCGGAAACACGGTTAGAA 9543
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9544 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGCAATCACGAACCAGTAAGGG 9603
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db 9604 TCTATGCCCAAGAGTCTGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9663
Qy 4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db 9664 ATTACTATCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTGCTGTGGCGA 9723
Qy 4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db 9724 TGATGATTGGCGTAACTGTTGCAGCATTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9783
Qy 4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db 9784 CGCCATATGCCCTGGCCCCAAATGCCGTGATTCCAACTTCGCTGGCACTTTTGTGCTGTG 9843
Qy 4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db 9844 TTAGGTCGGCTAATGCTGAAACATTCACCGAGACCATGAGTTACTTATGGTCGAACAGCC 9903
Qy 4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
Db 9904 AGCCGTTCTTCTGGGTCCAGCTGTGTATACCTCTGGCCGCTGTCGTCGTTCTAATGCGCT 9963

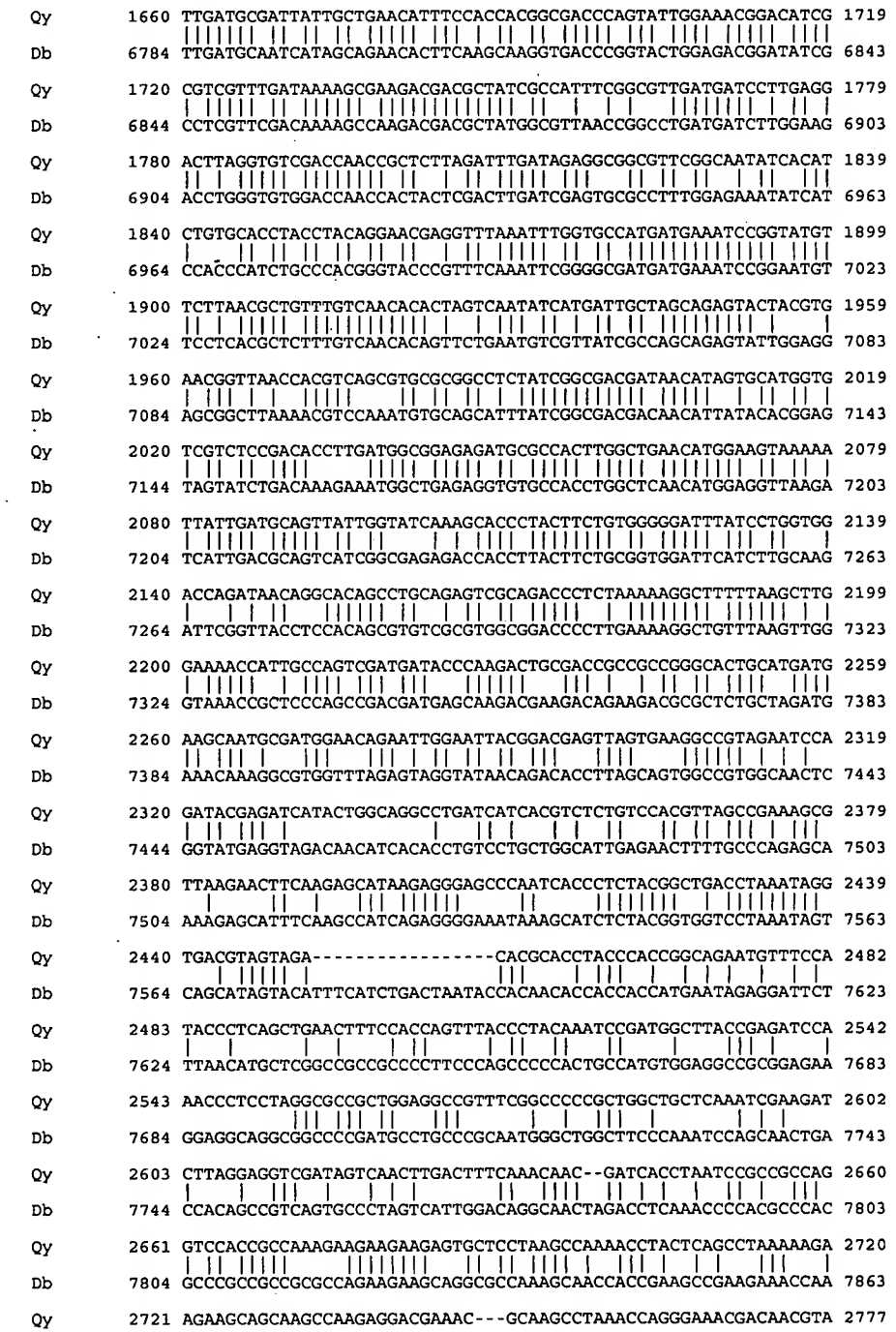
```
Qy       4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
Db       9964 GTTGCTCATGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10023

Qy       4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
Db       10024 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10083

Qy       4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db       10084 AAAGGGCAGGGTACGCCCCGCTCAATTTGGAGATTACTGTCATGTCCTCGGAGGTTTTGC 10143

Qy       4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db       10144 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCTAAAG 10203

Qy       5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db       10204 TCAGATGCTGCGGCTCCTTGGAATGTCAGCCCGCCGCTCACGCAGACTATACCTGCAAGG 10263

Qy       5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db       10264 TCTTTGGAGGGGTGTACCCCTTCATGTGGGGAGGAGCACAATGTTTTTGCGACAGTGAGA 10323

Qy       5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db       10324 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGTAGATTGCGCGACTGACCACGCGC 10383

Qy       5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db       10384 AGGCGATTAAGGTGCATACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10443

Qy       5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db       10444 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACCTGA 10503

Qy       5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db       10504 AAGTCATAGCTGGACCAATTTCAGCATTGTTTACACCATTCGATCACAAGGTCGTTATCA 10563

Qy       5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db       10564 ATCGCGGCCTGGTGTACAACTATGACTTTCCGGAATACGGAGCGATGAAACCAGGAGCGT 10623

Qy       5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db       10624 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAAGACCTCATCGCCAGCACAGACATTA 10683

Qy       5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db       10684 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCTGGAT 10743

Qy       5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db       10744 TCGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCCCCTTTTGGGTGCA 10803

Qy       5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db       10804 AGATTGCAGTCAATCCGCTTCGAGCGGTGGACTGCTCATACGGGAACATTCCCATTTCTA 10863

Qy       5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db       10864 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10923

Qy       5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db       10924 AATGTGATGTCAGTGAGTGCACTTATTCAGCGGACTTCGGAGGGATGGCTACCCTGCAGT 10983

Qy       5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db       10984 ATGTATCCGACCGCGAAGGACAATGCCCTGTACATTCGCATTCGAGCACAGCAACCCTCC 11043

Qy       5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942

Db       11044 AAGAGTCGACAGTTCATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTCAGCACCGCGA 11103

Qy       5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db       11104 GCCCACAGGCGAACTTCATTGTATCGCTGTGTGGTAAGAAGACAACATGCAATGCAGAAT 11163

Qy       6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db       11164 GCAAACCACCAGCTGATCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTCCAAG 11223

Qy       6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db       11224 CCGCCATCTCAAAAACTTCATGGAGTTGGCTGTTTGCCCTTTCGGCGGCGCCTCGTCGC 11283

Qy       6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
Db       11284 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11343

Qy       6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
Db       11344 GACCGCTACGCCCCAATGACCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11403

Qy       6242 GCATAATGCCACGCGCCGCT----------TGACACT 6268
Db       11404 GCATAATGCATCAGGCTGGTATATTAGATCCCCGCTTACCGCGGGCAATATAGCAACACC 11463

Qy       6269 AAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGT 6318
Db       11464 AAAACTCGACGTATTTCCGAGGAAGCGCAGTGCATAATGCTGCGCAGTGT 11513
```

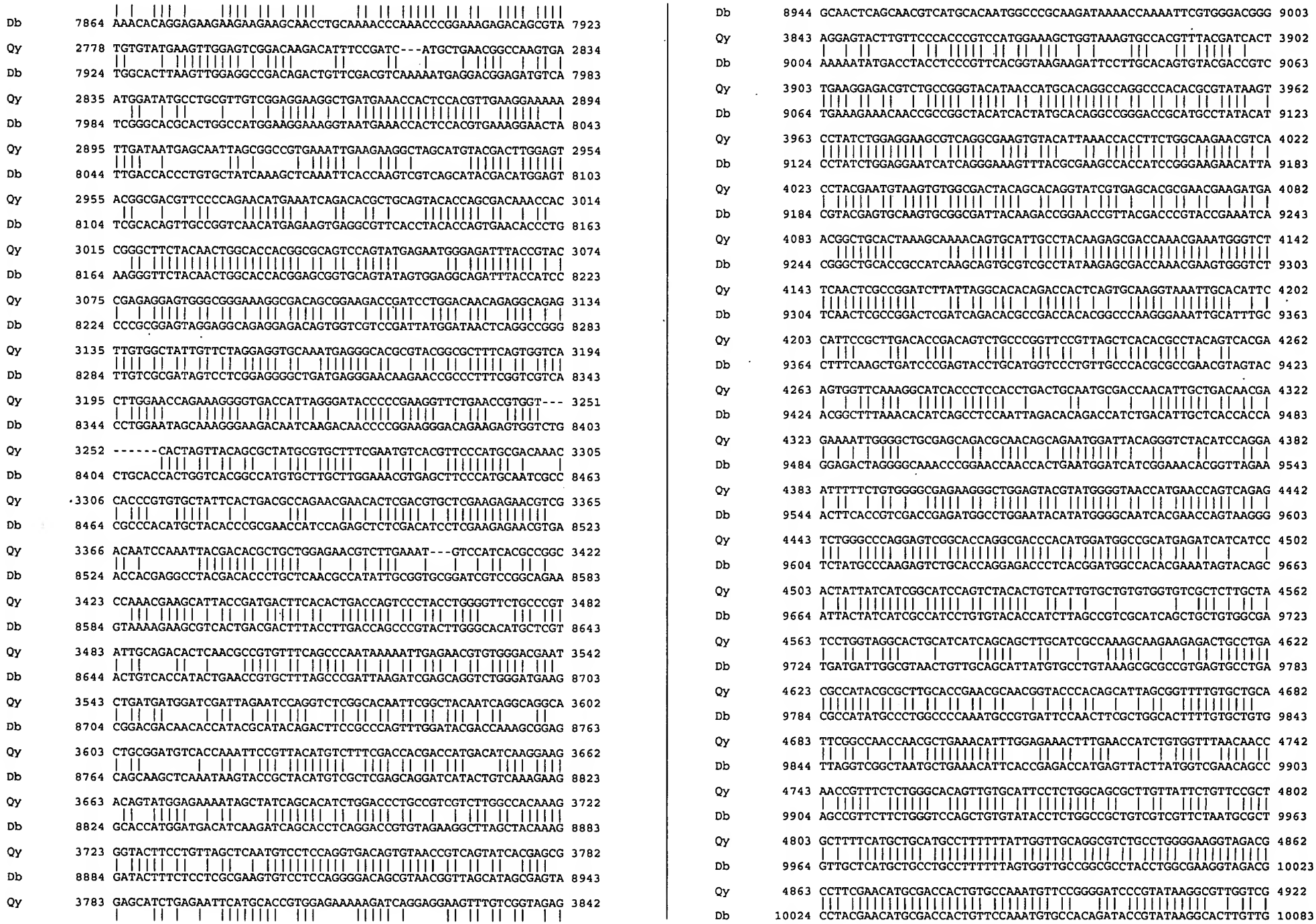
RESULT 10
BD071120

```
LOCUS       BD071120               11663 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  System for the in vivo delivery and expression of heterologous
            genes in the bone marrow.
ACCESSION   BD071120
VERSION     BD071120.1  GI:22616723
KEYWORDS    JP 2001515348-A/1.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 11663)
  AUTHORS   Johnston,R.E., Davis,N.L. and Simpson,D.A.
  TITLE     System for the in vivo delivery and expression of heterologous
            genes in the bone marrow
  JOURNAL   Patent: JP 2001515348-A.1 18-SEP-2001;
            UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
COMMENT     OS   Girdwood S.A.
            PN   JP 2001515348-A/1
            PD   18-SEP-2001
            PF   18-FEB-1998 JP 1998536760
            PR   19-FEB-1997 US    08/801263
            PI   ROBERT E JOHNSTON,NANCY L DAVIS,DENNIS A SIMPSON PC
            C12N15/86,C12N15/33,C12N7/01,C12N5/10,A61K39/12,A61K48/00 CC
            System for the in vivo delivery and expression of heterologous
            CC            genes in
            CC   the bone marrow
            FH   Key             Location/Qualifiers
            FT   CDS             (60). .(7559)
            FT   CDS             (7608). .(11342).
FEATURES             Location/Qualifiers
     source          1. .11663
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT     3291 a   3105 c   2907 g   2360 t
ORIGIN
```

```
  Query Match           36.9%;  Score 2396;  DB 6;  Length 11663;
  Best Local Similarity   64.6%;  Pred. No. 0;
  Matches 3751;  Conservative    0;  Mismatches 1985;  Indels   74;  Gaps    9;
```





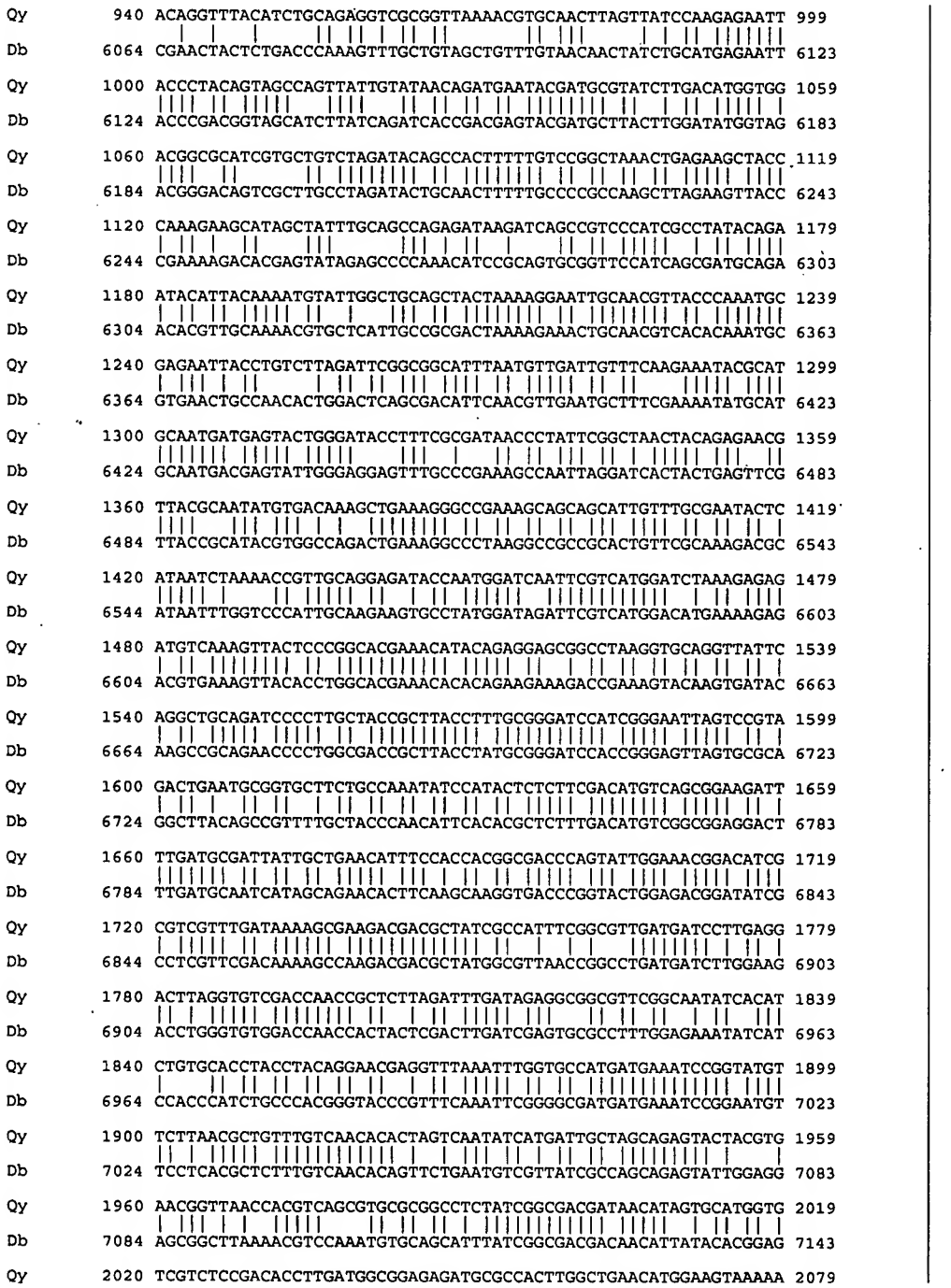


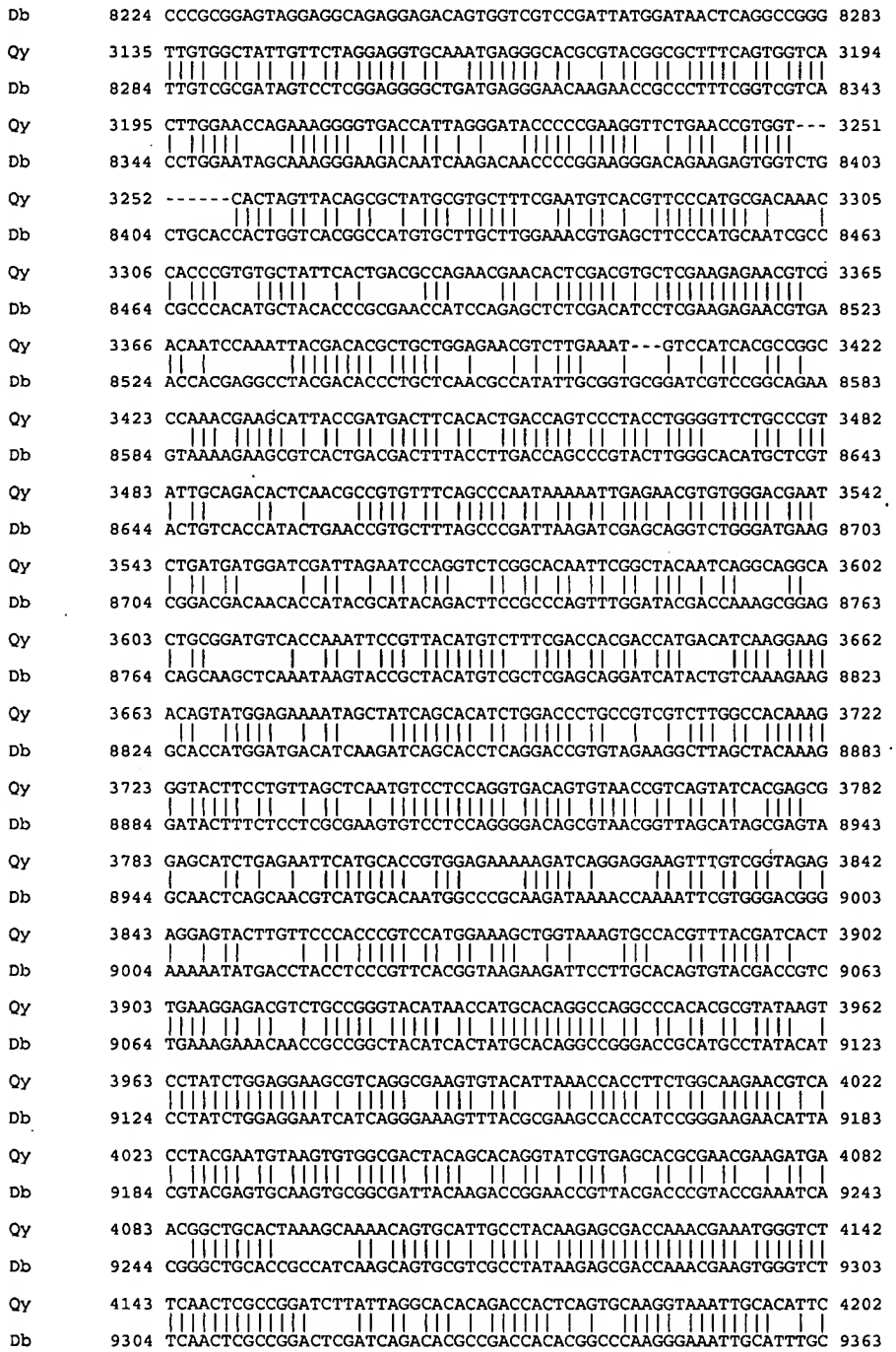
```
Qy       4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db      10084 AAAGGGCAGGGTACGCCCCGCTCAATTTGGAGATTACTGTCATGTCCTCGGAGGTTTTGC 10143

Qy       4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db      10144 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCTAAAG 10203

Qy       5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db      10204 TCAGATGCTGCGGCTCCTTGGAATGTCAGCCCGCCGCTCACGCAGACTATACCTGCAAGG 10263

Qy       5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db      10264 TCTTTGGAGGGGTGTACCCCTTCATGTGGGGAGGAGCACAATGTTTTTGCGACAGTGAGA 10323

Qy       5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db      10324 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGTAGATTGCGCGACTGACCACGCGC 10383

Qy       5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db      10384 AGGCGATTAAGGTGCATACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10443

Qy       5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db      10444 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACCTGA 10503

Qy       5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db      10504 AAGTCATAGCTGGACCAATTTCAGCATTGTTTACACCATTCGATCACAAGGTCGTTATCA 10563

Qy       5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db      10564 ATCGCGGCCTGGTGTACAACTATGACTTTCCGGAATACGGAGCGATGAAACCAGGAGCGT 10623

Qy       5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db      10624 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAAGACCTCATCGCCAGCACAGACATTA 10683

Qy       5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db      10684 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCTGGAT 10743

Qy       5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db      10744 TCGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCCCCTTTTGGGTGCA 10803

Qy       5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db      10804 AGATTGCAGTCAATCCGCTTCGAGCGGTGGACTGCTCATACGGGAACATTCCCATTTCTA 10863

Qy       5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db      10864 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10923

Qy       5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db      10924 AATGTGATGTCAGTGAGTGCACTTATTCAGCGGACTTCGGAGGGATGGCTACCCTGCAGT 10983

Qy       5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db      10984 ATGTATCCGACCGCGAAGGACAATGCCCTGTACATTCGCATTCGAGCACAGCAACCCTCC 11043

Qy       5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db      11044 AAGAGTCGACAGTTCATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTCAGCACCGCGA 11103

Qy       5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db      11104 GCCCACAGGCGAACTTCATTGTATCGCTGTGTGGTAAGAAGACAACATGCAATGCAGAAT 11163

Qy       6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db      11164 GCAAACCACCAGCTGATCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTCCAAG 11223

Qy       6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db      11224 CCGCCATCTCAAAAACTTCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11283

Qy       6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
Db      11284 TATTAATTATAGGACTTATGATTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11343

Qy       6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
Db      11344 GACCGCTACGCCCCAATGACCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11403

Qy       6242 GCATAATGCCACGCGCCGCT----------------------------------TGACACT 6268
Db      11404 GCATAATGCATCAGGCTGGTATATTAGATCCCCGCTTACCGCGGGCAATATAGCAACACC 11463

Qy       6269 AAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGT 6318
Db      11464 AAAACTCGACGTATTTCCGAGGAAGCGCAGTGCATAATGCTGCGCAGTGT 11513
```

```
RESULT 11
I46902
LOCUS       I46902                 11663 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 1 from patent US 5639650.
ACCESSION   I46902
VERSION     I46902.1  GI:2470867
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 11663)
  AUTHORS   Johnston,R.E., Davis,N.L. and Simpson,D.A.
  TITLE     cDNA clone for South African Arbovirus No. 86
  JOURNAL   Patent: US 5639650-A 1 17-JUN-1997;
FEATURES             Location/Qualifiers
     source          1. .11663
                     /organism="unknown"
BASE COUNT     3291 a   3105 c   2907 g   2360 t
ORIGIN
```

Query Match 36.9%; Score 2396; DB 6; Length 11663;
Best Local Similarity 64.6%; Pred. No. 0;
Matches 3751; Conservative 0; Mismatches 1985; Indels 74; Gaps 9;

```
Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
Db       5707 ACTGTCTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5766

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
Db       5767 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTTCTGG 5826

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
Db       5827 AAAGAATCTACGCCCCGGTGCTCGACACGTCGAAAGAGGAACAGCTCAAACTCAGGTACC 5886

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
Db       5887 AGATGATGCCCACCGAAGCCAACAAAAGCAGGTACCAGTCTCGAAAAGTAGAAAACCAGA 5946

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
Db       5947 AAGCCATAACCACTGAGCGACTGCTTTCAGGGCTACGACTGTATAACTCTGCCACAG---6003

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
Db       6004 ATCAGCCAGAATGCTATAAGATCACCTACCCGAAACCATCGTATTCCAGCAGTGTACCAG 6063
```

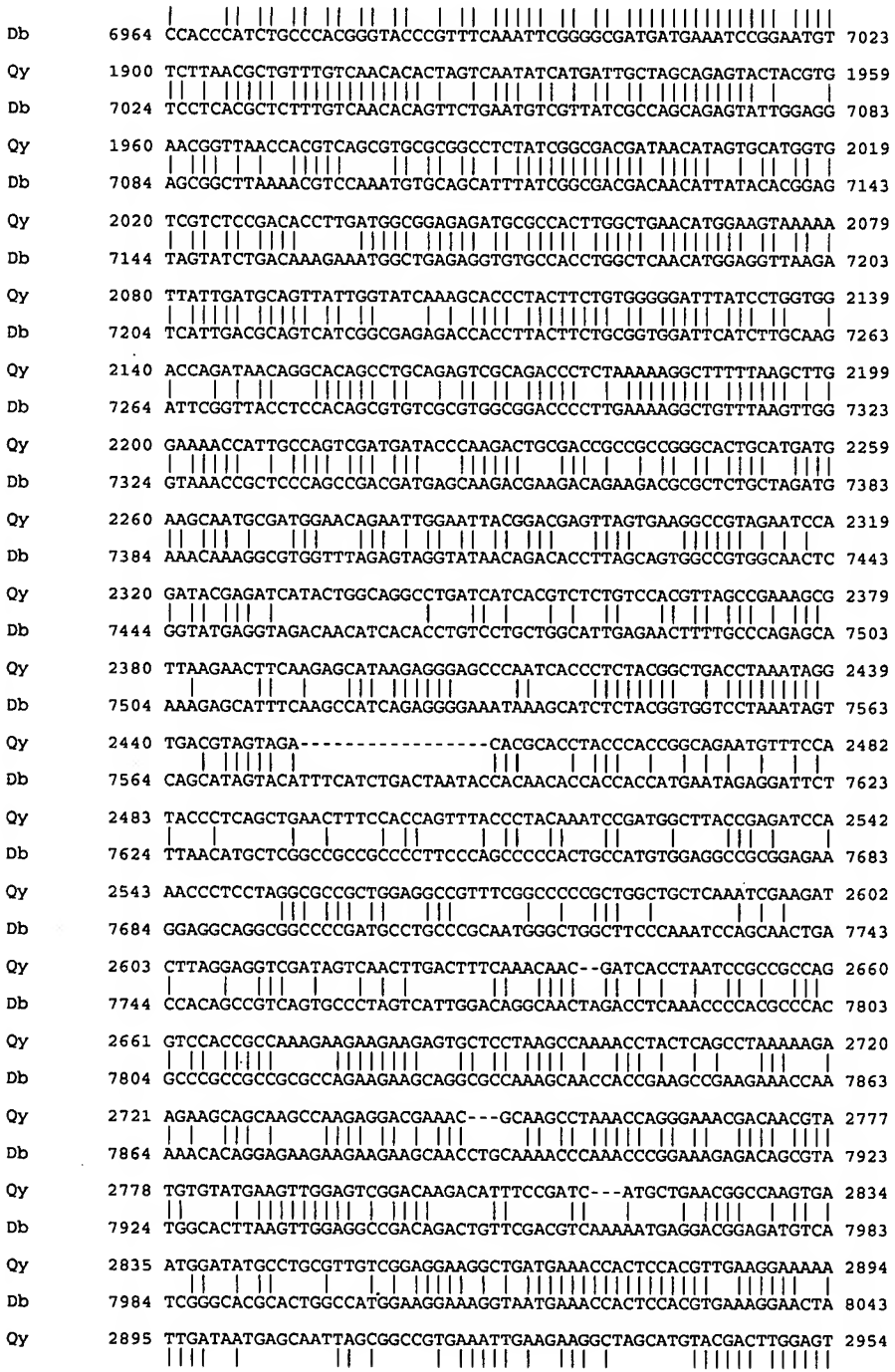

Qy 940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
Db 6064 CGAACTACTCTGACCCAAAGTTTGCTGTAGCTGTTTGTAACAACTATCTGCATGAGAATT 6123
Qy 1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
Db 6124 ACCCGACGGTAGCATCTTATCAGATCACCGACGAGTACGATGCTTACTTGGATATGGTAG 6183
Qy 1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
Db 6184 ACGGGACAGTCGCTTGCCTAGATACTGCAACTTTTTGCCCCGCCAAGCTTAGAAGTTACC 6243
Qy 1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
Db 6244 CGAAAAGACACGAGTATAGAGCCCCAAACATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6303
Qy 1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
Db 6304 ACACGTTGCAAAACGTGCTCATTGCCGCGACTAAAAGAAACTGCAACGTCACACAAATGC 6363
Qy 1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
Db 6364 GTGAACTGCCAACACTGGACTCAGCGACATTCAACGTTGAATGCTTTCGAAAATATGCAT 6423
Qy 1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
Db 6424 GCAATGACGAGTATTGGGAGGAGTTTGCCCGAAAGCCAATTAGGATCACTACTGAGTTCG 6483
Qy 1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
Db 6484 TTACCGCATACGTGGCCAGACTGAAAGGCCCTAAGGCCGCCGCACTGTTCGCAAAGACGC 6543
Qy 1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
Db 6544 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6603
Qy 1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
Db 6604 ACGTGAAAGTTACACCTGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6663
Qy 1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
Db 6664 AAGCCGCAGAACCCCTGGCGACCGCTTACCTATGCGGGATCCACCGGGAGTTAGTGCGCA 6723
Qy 1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
Db 6724 GGCTTACAGCCGTTTTGCTACCCAACATTCACACGCTCTTTGACATGTCGGCGGAGGACT 6783
Qy 1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
Db 6784 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGTGACCCGGTACTGGAGACGGATATCG 6843
Qy 1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
Db 6844 CCTCGTTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGCCTGATGATCTTGGAAG 6903
Qy 1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
Db 6904 ACCTGGGTGTGGACCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 6963
Qy 1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
Db 6964 CCACCCATCTGCCCACGGGTACCCGTTTCAAATTCGGGGCGATGATGAAATCCGGAATGT 7023
Qy 1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
Db 7024 TCCTCACGCTCTTTGTCAACACAGTTCTGAATGTCGTTATCGCCAGCAGAGTATTGGAGG 7083
Qy 1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
Db 7084 AGCGGCTTAAAACGTCCAAATGTGCAGCATTTATCGGCGACGACAACATTATACACGGAG 7143
Qy 2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
Db 7144 TAGTATCTGACAAAGAAATGGCTGAGAGGTGTGCCACCTGGCTCAACATGGAGGTTAAGA 7203
Qy 2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139
Db 7204 TCATTGACGCAGTCATCGGCGAGAGACCACCTTACTTCTGCGGTGGATTCATCTTGCAAG 7263
Qy 2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199
Db 7264 ATTCGGTTACCTCCACAGCGTGTCGCGTGGCGGACCCCTTGAAAAGGCTGTTTAAGTTGG 7323
Qy 2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
Db 7324 GTAAACCGCTCCCAGCCGACGATGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 7383
Qy 2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319
Db 7384 AAACAAAGGCGTGGTTTAGAGTAGGTATAACAGACACCTTAGCAGTGGCCGTGGCAACTC 7443
Qy 2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379
Db 7444 GGTATGAGGTAGACAACATCACACCTGTCCTGCTGGCATTGAGAACTTTTGCCCAGAGCA 7503
Qy 2380 TTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGG 2439
Db 7504 AAAGAGCATTTCAAGCCATCAGAGGGGAAATAAAGCATCTCTACGGTGGTCCTAAATAGT 7563
Qy 2440 TGACGTAGTAGA----------CACGCACCTACCCACCGGCAGAATGTTTCCA 2482
Db 7564 CAGCATAGTACATTTCATCTGACTAATACCACAACACCACCACCATGAATAGAGGATTCT 7623
Qy 2483 TACCCTCAGCTGAACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCA 2542
Db 7624 TTAACATGCTCGGCCGCCGCCCCTTCCCAGCCCCCACTGCCATGTGGAGGCCGCGGAGAA 7683
Qy 2543 AACCCTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGAT 2602
Db 7684 GGAGGCAGGCGGCCCCGATGCCTGCCCGCAATGGGCTGGCTTCCCAAATCCAGCAACTGA 7743
Qy 2603 CTTAGGAGGTCGATAGTCAACTTGACTTTCAAACAAC--GATCACCTAATCCGCCGCCAG 2660
Db 7744 CCACAGCCGTCAGTGCCCTAGTCATTGGACAGGCAACTAGACCTCAAACCCCACGCCCAC 7803
Qy 2661 GTCCACCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGA 2720
Db 7804 GCCCGCCGCCGCGCCAGAAGAAGCAGGCGCCAAAGCAACCACCGAAGCCGAAGAAACCAA 7863
Qy 2721 AGAAGCAGCAAGCCAAGAGGACGAAAC---GCAAGCCTAAACCAGGGAAACGACAACGTA 2777
Db 7864 AAACACAGGAGAAGAAGAAGAAGCAACCTGCAAAACCCAAACCCGGAAAGAGACAGCGTA 7923
Qy 2778 TGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATC---ATGCTGAACGGCCAAGTGA 2834
Db 7924 TGGCACTTAAGTTGGAGGCCGACAGACTGTTCGACGTCAAAAATGAGGACGGAGATGTCA 7983
Qy 2835 ATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACTCCACGTTGAAGGAAAAA 2894
Db 7984 TCGGGCACGCACTGGCCATGGAAGGAAAGGTAATGAAACCACTCCACGTGAAAGGAACTA 8043
Qy 2895 TTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAGCATGTACGACTTGGAGT 2954
Db 8044 TTGACCACCCTGTGCTATCAAAGCTCAAATTCACCAAGTCGTCAGCATACGACATGGAGT 8103
Qy 2955 ACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTACACCAGCGACAAACCAC 3014
Db 8104 TCGCACAGTTGCCGGTCAACATGAGAAGTGAGGCGTTCACCTACACCAGTGAACACCCTG 8163
Qy 3015 CGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAATGGGAGATTTACCGTAC 3074
Db 8164 AAGGGTTCTACAACTGGCACCACGGAGCGGTGCAGTATAGTGGAGGCAGATTTACCATCC 8223
Qy 3075 CGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGACAACAGAGGCAGAG 3134

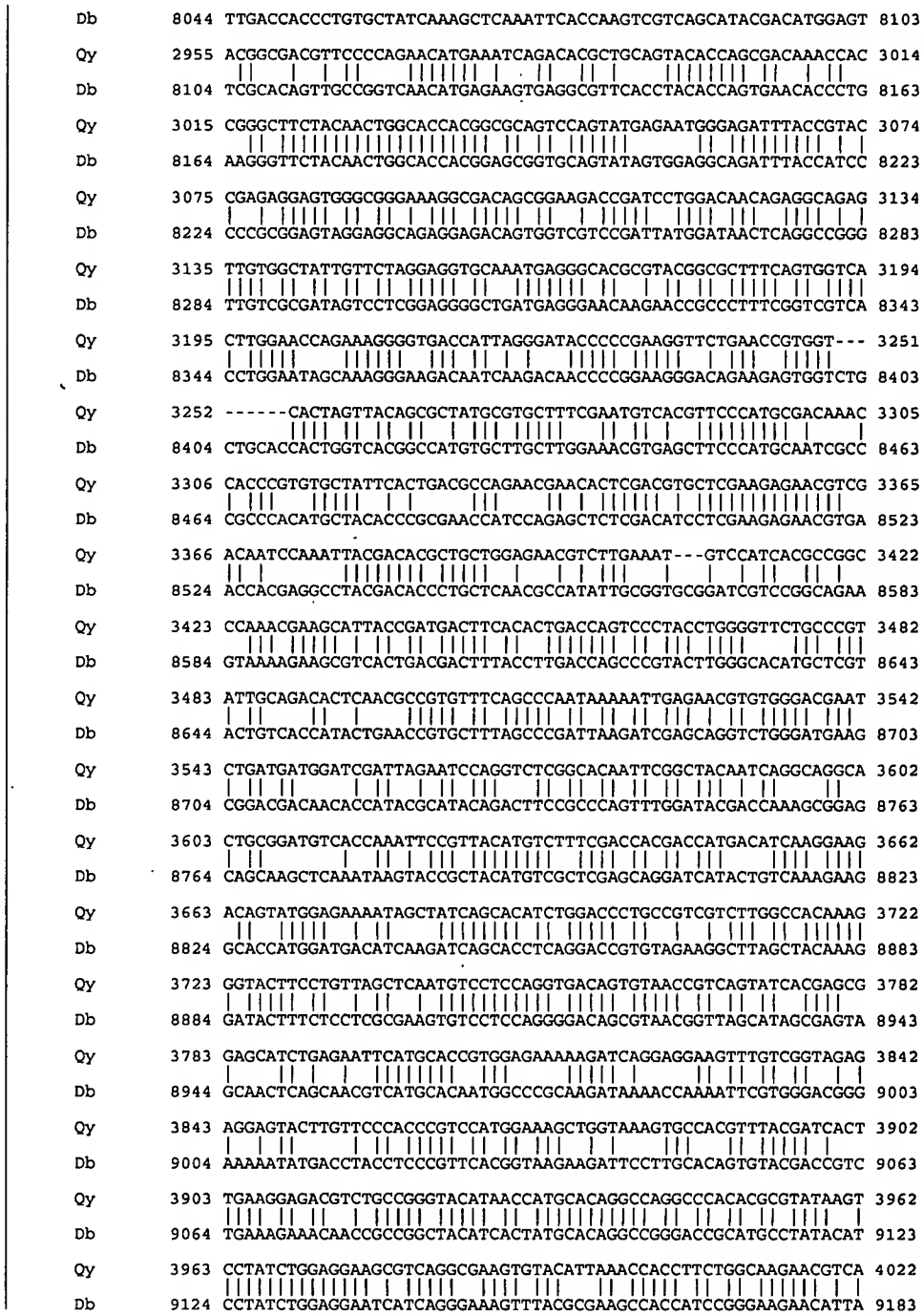

Db 8224 CCCGCGGAGTAGGAGGCAGAGGAGACAGTGGTCGTCCGATTATGGATAACTCAGGCCGGG 8283
Qy 3135 TTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGCGCTTTCAGTGGTCA 3194
Db 8284 TTGTCGCGATAGTCCTCGGAGGGGCTGATGAGGGAACAAGAACCGCCCTTTCGGTCGTCA 8343
Qy 3195 CTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTCTGAACCGTGGT--- 3251
Db 8344 CCTGGAATAGCAAAGGGAAGACAATCAAGACAACCCCGGAAGGGACAGAAGAGTGGTCTG 8403
Qy 3252 ------CACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGACAAAC 3305
Db 8404 CTGCACCACTGGTCACGGCCATGTGCTTGCTTGGAAACGTGAGCTTCCCATGCAATCGCC 8463
Qy 3306 CACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAACGTCG 3365
Db 8464 CGCCCACATGCTACACCCGCGAACCATCCAGAGCTCTCGACATCCTCGAAGAGAACGTGA 8523
Qy 3366 ACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAAT---GTCCATCACGCCGGC 3422
Db 8524 ACCACGAGGCCTACGACACCCTGCTCAACGCCATATTGCGGTGCGGATCGTCCGGCAGAA 8583
Qy 3423 CCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCCCGT 3482
Db 8584 GTAAAAGAAGCGTCACTGACGACTTTACCTTGACCAGCCCGTACTTGGGCACATGCTCGT 8643
Qy 3483 ATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACGAAT 3542
Db 8644 ACTGTCACCATACTGAACCGTGCTTTAGCCCGATTAAGATCGAGCAGGTCTGGGATGAAG 8703
Qy 3543 CTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAGGCA 3602
Db 8704 CGGACGACAACACCATACGCATACAGACTTCCGCCCAGTTTGGATACGACCAAAGCGGAG 8763
Qy 3603 CTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAAGGAAG 3662
Db 8764 CAGCAAGCTCAAATAAGTACCGCTACATGTCGCTCGAGCAGGATCATACTGTCAAAGAAG 8823
Qy 3663 ACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCACAAAG 3722
Db 8824 GCACCATGGATGACATCAAGATCAGCACCTCAGGACCGTGTAGAAGGCTTAGCTACAAAG 8883
Qy 3723 GGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCACGAGCG 3782
Db 8884 GATACTTTCTCCTCGCGAAGTGTCCTCCAGGGGACAGCGTAACGGTTAGCATAGCGAGTA 8943
Qy 3783 GAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTAGAG 3842
Db 8944 GCAACTCAGCAACGTCATGCACAATGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGG 9003
Qy 3843 AGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATCACT 3902
Db 9004 AAAAATATGACCTACCTCCCGTTCACGGTAAGAAGATTCCTTGCACAGTGTACGACCGTC 9063
Qy 3903 TGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAAGT 3962
Db 9064 TGAAAGAAACAACCGCCGGCTACATCACTATGCACAGGCCGGGACCGCATGCCTATACAT 9123
Qy 3963 CCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGTCA 4022
Db 9124 CCTATCTGGAGGAATCATCAGGGAAAGTTTACGCGAAGCCACCATCCGGGAAGAACATTA 9183
Qy 4023 CCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGATGA 4082
Db 9184 CGTACGAGTGCAAGTGCGGCGATTACAAGACCGGAACCGTTACGACCCGTACCGAAATCA 9243
Qy 4083 ACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGTCT 4142
Db 9244 CGGGCTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAACGAAGTGGGTCT 9303
Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9304 TCAACTCGCCGGACTCGATCAGACACGCCGACCACACGGCCCAAGGGAAATTGCATTTGC 9363
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9364 CTTTCAAGCTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAACGTAGTAC 9423
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9424 ACGGCTTTAAACACATCAGCCTCCAATTAGACACAGACCATCTGACATTGCTCACCACCA 9483
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9484 GGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCATCGGAAACACGGTTAGAA 9543
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9544 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGCAATCACGAACCAGTAAGGG 9603
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db 9604 TCTATGCCCAAGAGTCTGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9663
Qy 4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db 9664 ATTACTATCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTGCTGTGGCGA 9723
Qy 4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db 9724 TGATGATTGGCGTAACTGTTGCAGCATTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9783
Qy 4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db 9784 CGCCATATGCCCTGGCCCCAAATGCCGTGATTCCAACTTCGCTGGCACTTTTGTGCTGTG 9843
Qy 4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db 9844 TTAGGTCGGCTAATGCTGAAACATTCACCGAGACCATGAGTTACTTATGGTCGAACAGCC 9903
Qy 4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
Db 9904 AGCCGTTCTTCTGGGTCCAGCTGTGTATACCTCTGGCCGCTGTCGTCGTTCTAATGCGCT 9963
Qy 4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
Db 9964 GTTGCTCATGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10023
Qy 4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
Db 10024 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10083
Qy 4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db 10084 AAAGGGCAGGGTACGCCCCGCTCAATTTGGAGATTACTGTCATGTCCTCGGAGGTTTTGC 10143
Qy 4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db 10144 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCTAAAG 10203
Qy 5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db 10204 TCAGATGCTGCGGCTCCTTGGAATGTCAGCCCGCCGCTCACGCAGACTATACCTGCAAGG 10263
Qy 5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db 10264 TCTTTGGAGGGGTGTACCCCTTCATGTGGGGAGGAGCACAATGTTTTTGCGACAGTGAGA 10323
Qy 5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db 10324 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGTAGATTGCGCGACTGACCACGCGC 10383
Qy 5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db 10384 AGGCGATTAAGGTGCATACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10443

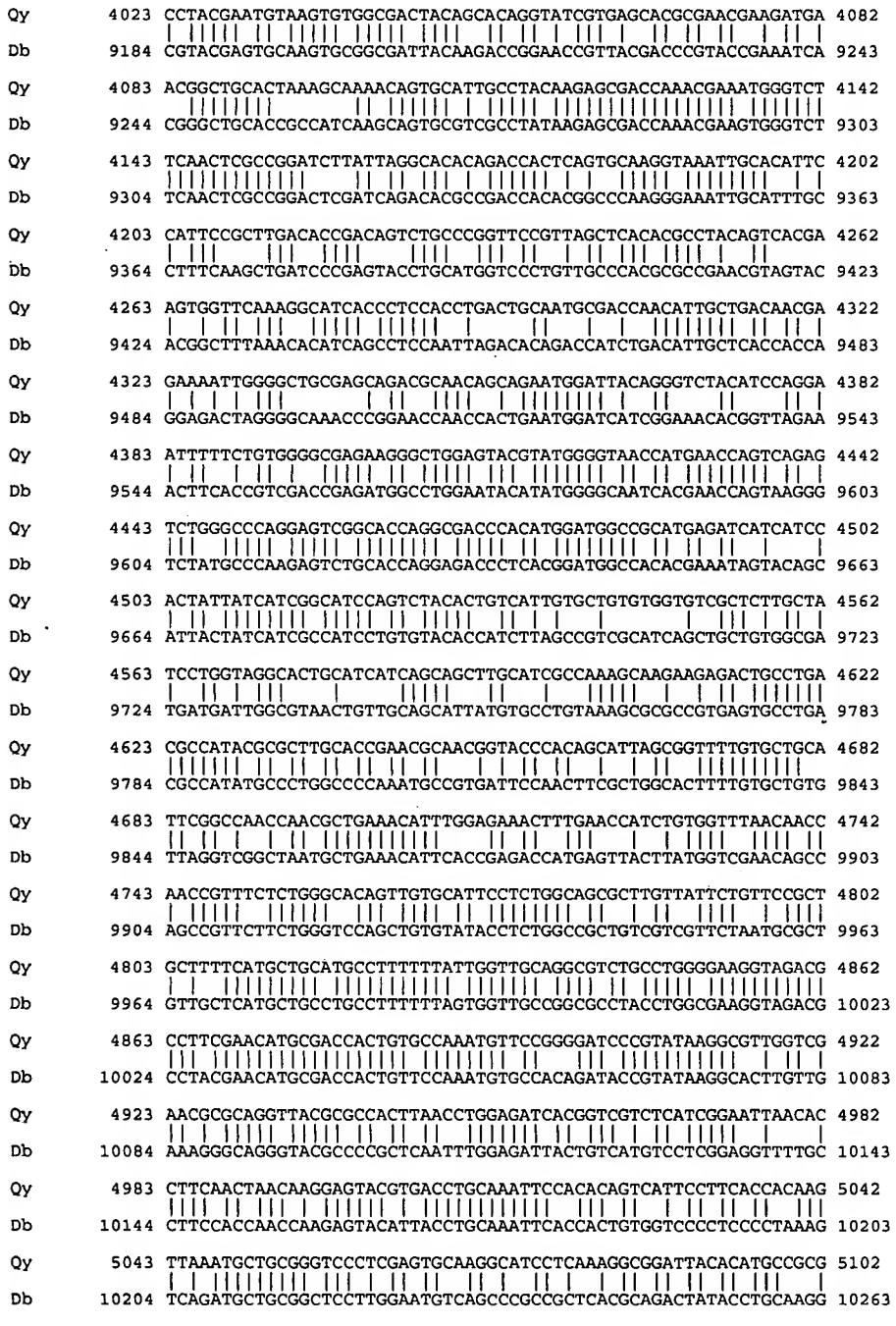
```
Qy       5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db      10444 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACCTGA 10503

Qy       5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db      10504 AAGTCATAGCTGGACCAATTTCAGCATTGTTTACACCATTCGATCACAAGGTCGTTATCA 10563

Qy       5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db      10564 ATCGCGGCCTGGTGTACAACTATGACTTTCCGGAATACGGAGCGATGAAACCAGGAGCGT 10623

Qy       5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db      10624 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAAGACCTCATCGCCAGCACAGACATTA 10683

Qy       5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db      10684 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCTGGAT 10743

Qy       5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db      10744 TCGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCCCCTTTTGGGTGCA 10803

Qy       5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db      10804 AGATTGCAGTCAATCCGCTTCGAGCGGTGGACTGCTCATACGGGAACATTCCCATTTCTA 10863

Qy       5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db      10864 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10923

Qy       5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db      10924 AATGTGATGTCAGTGAGTGCACTTATTCAGCGGACTTCGGAGGGATGGCTACCCTGCAGT 10983

Qy       5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db      10984 ATGTATCCGACCGCGAAGGACAATGCCCTGTACATTCGCATTCGAGCACAGCAACCCTCC 11043

Qy       5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db      11044 AAGAGTCGACAGTTCATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTCAGCACCGCGA 11103

Qy       5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db      11104 GCCCACAGGCGAACTTCATTGTATCGCTGTGTGGTAAGAAGACAACATGCAATGCAGAAT 11163

Qy       6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db      11164 GCAAACCACCAGCTGATCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTCCAAG 11223

Qy       6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db      11224 CCGCCATCTCAAAAACTTCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11283

Qy       6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
Db      11284 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11343

Qy       6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
Db      11344 GACCGCTACGCCCCAATGACCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11403

Qy       6242 GCATAATGCCACGCGCCGCT----------------------------------TGACACT 6268
Db      11404 GCATAATGCATCAGGCTGGTATATTAGATCCCCGCTTACCGCGGGCAATATAGCAACACC 11463

Qy       6269 AAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGT 6318
Db      11464 AAAACTCGACGTATTTCCGAGGAAGCGCAGTGCATAATGCTGCGCAGTGT 11513
```

```
RESULT 12
ACU38305
LOCUS       ACU38305               11663 bp    RNA     linear   VRL 30-JAN-1996
DEFINITION  Sindbis-like virus isolate S.A.AR86, complete genome.
ACCESSION   U38305
VERSION     U38305.1  GI:1125069
KEYWORDS    .
SOURCE      Sindbis-like virus
  ORGANISM  Sindbis-like virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
            Alphavirus; WEEV complex.
REFERENCE   1  (sites)
  AUTHORS   Russell,D.L., Dalrymple,J.M. and Johnston,R.E.
  TITLE     Sindbis virus mutations which coordinately affect glycoprotein
            processing, penetration, and virulence in mice
  JOURNAL   J. Virol. 63 (4), 1619-1629 (1989)
  MEDLINE   89178848
   PUBMED   2926866
REFERENCE   2  (bases 1 to 11663)
  AUTHORS   Simpson,D.A., Lin,S.C. and Johnston,R.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-1995) Dennis A. Simpson, Microbiology &
            Immunology, University of North Carolina, Rm834 FLOB, CB7290,
            Chapel Hill, NC 27599-7290, USA
FEATURES             Location/Qualifiers
     source          1. .11663
                     /organism="Sindbis-like virus"
                     /mol_type="genomic RNA"
                     /isolate="S.A.AR86"
                     /db_xref="taxon:44419"
                     /note="from South Africa."
     5'UTR           1. .59
     CDS             60. .7562
                     /codon_start=1
                     /product="nonstructural polyprotein"
                     /protein_id="AAA86135.1"
                     /db_xref="GI:1125070"
                     /translation="MEKPVVNVDVDPQSPFVVQLQKSFPQFEVVAQQVTPNDHANARA
                     FSHLASKLIELEVPTTATILDIGSAPARRMFSEHQYHCVCPMRSPEDPDRMMKYASKL
                     AEKACKITNKNLHEKIKDLRTVLDTPDAETPSLCFHNDVTCNTRAEYSVMQDVYINAP
                     GTIYHQAMKGVRTLYWIGFDTTQFMFSAMAGSYPAYNTNWADEKVLEARNIGLCSTKL
                     SEGRTGKLSIMRKKELKPGSRVYFSVGSTLYPEHRASLQSWHLPSVFHLKGKQSYTCR
                     CDTVVSCEGYVVKKITISPGITGETVGYAVTNNSEGFLLCKVTDTVKGERVSFPVCTY
                     IPATICDQMTGIMATDISPDDAQKLLVGLNQRIVINGKTNRNTNTMQNYLLPIIAQGF
                     SKWAKERKEDLDNEKMLGTRERKLTYGCLWAFRTKKVHSFYRPPGTQTIVKVPASFSA
                     FPMSSVWTTSLPMSLRQKMKLALQPKKEEKLLQVPEELVMEAKAAFEDAQEESRAEKL
                     REALPPLVADKGIEAAAEVVCEVEGLQADTGAALVETPRGHVRIIPQANDRMIGQYIV
                     VSPISVLKNAKLAPAHPLADQVKIITHSGRSGRYAVEPYDAKVLMPAGSAVPWPEFLA
                     LSESATLVYNEREFVNRKLYHIAMHGPAKNTEEEQYKVTKAELAETEYVFDVDKKRCV
                     KKEEASGLVLSGELTNPPYHELALEGLKTRPAVPYKVETIGVIGTPGSGKSAIIKSTV
                     TARDLVTSGKKENCREIEADVLRLRGMQITSKTVDSVMLNGCHKAVEVLYVDEAFRCH
                     AGALLALIAIVRPRKKVVLCGDPKQCGFFNMMQLKVHFNHPEKDICTKTFYKFISRRC
                     TQPVTAIVSTLHYDGKMKTTNPCKKNIEIDITGATKPKPGDIILTCFRGWVKQLQIDY
                     PGHEVMTAAASQGLTRKGVYAVRQKVNENPLYAITSEHVNVLLTRTEDRLVWKTLQGD
                     PWIKQLTNVPKGNFQATIEDWEAEHKGIIAAINSPAPRTNPFSCKTNVCWAKALEPIL
                     ATAGIVLTGCQWSELFPQFADDKPHSAIYALDVICIKFFGMDLTSGLFSKQSIPLTYH
                     PADSARPVAHWDNSPGTRKYGYDHAVAAELSRRFPVFQLAGKGTQLDLQTGRTRVISA
                     QHNLVPVNRNLPHALVPEHKEKQPGPVEKFLSQFKHHSVLVISEKKIEAPHKRIEWIA
                     PIGIAGADKNYNLAFGFPPQARYDLVFINIGTKYRNHHFQQCEDHAATLKTLSRSALN
                     CLNPGGTLVVKSYGYADRNSEDVVTALARKFVRVSAARPECVSSNTEMYLIFRQLDNS
                     RTRQFTPHHLNCVISSVYEGTRDGVGAAPSYRTKRENIADCQEEAVVNAANPLGRPGE
                     GVCRAIYKRWPNSFTDSATETGTAKLTVCQGKKVIHAVGPDFRKHPEAEALKLLQNAY
                     HAVADLVNEHNIKSVAIPLLSTGIYAAGKDRLEVSLNCLLTALDRTDADVTIYCLDKK
                     WKERIDAVLQLKESVTELKDEDMEIDDELVWIHPDSCLKGRKGFSTTKGKLYSYFEGT
                     KFHQAAKDMAEIKVLFPNDQESNEQLCAYILGETMEAIREKCPVDHNPSSSPPKTLPC
                     LCMYAMTPERVHRLRSNNVKEVTVCSSTPLPKYKIKNVQKVQCTKVVLFNPHTPAFVP
                     ARKYIEAPEQPAAPPAQAEEAPGVVATPTPPAADNTSLDVTDISLDMEDSSEGSLFSS
                     FSGSDNYRRQVVVADVHAVQEPAPVPPPRLKKMARLAAARMQEEPTPPASTSSADESL
                     HLSFDGVSISFGSLFDGEMARLAAAQPPASTCPTDVPMSFGSFSDGEIEELSRRVTES
                     EPVLFGSFEPGEVNSIISSRSAVSFPPRKQRRRRRSRRTEYCLTGVGGYIFSTDTGPG
                     HLQKKSVLQNQLTEPTLERNVLERIYAPVLDTSKEEQLKLRYQMMPTEANKSRYQSRK
                     VENQKAITTERLLSGLRLYNSATDQPECYKITYPKPSYSSSVPANYSDPKFAVAVCNN
                     YLHENYPTVASYQITDEYDAYLDMVDGTVACLDTATFCPAKLRSYPKRHEYRAPNIRS
```

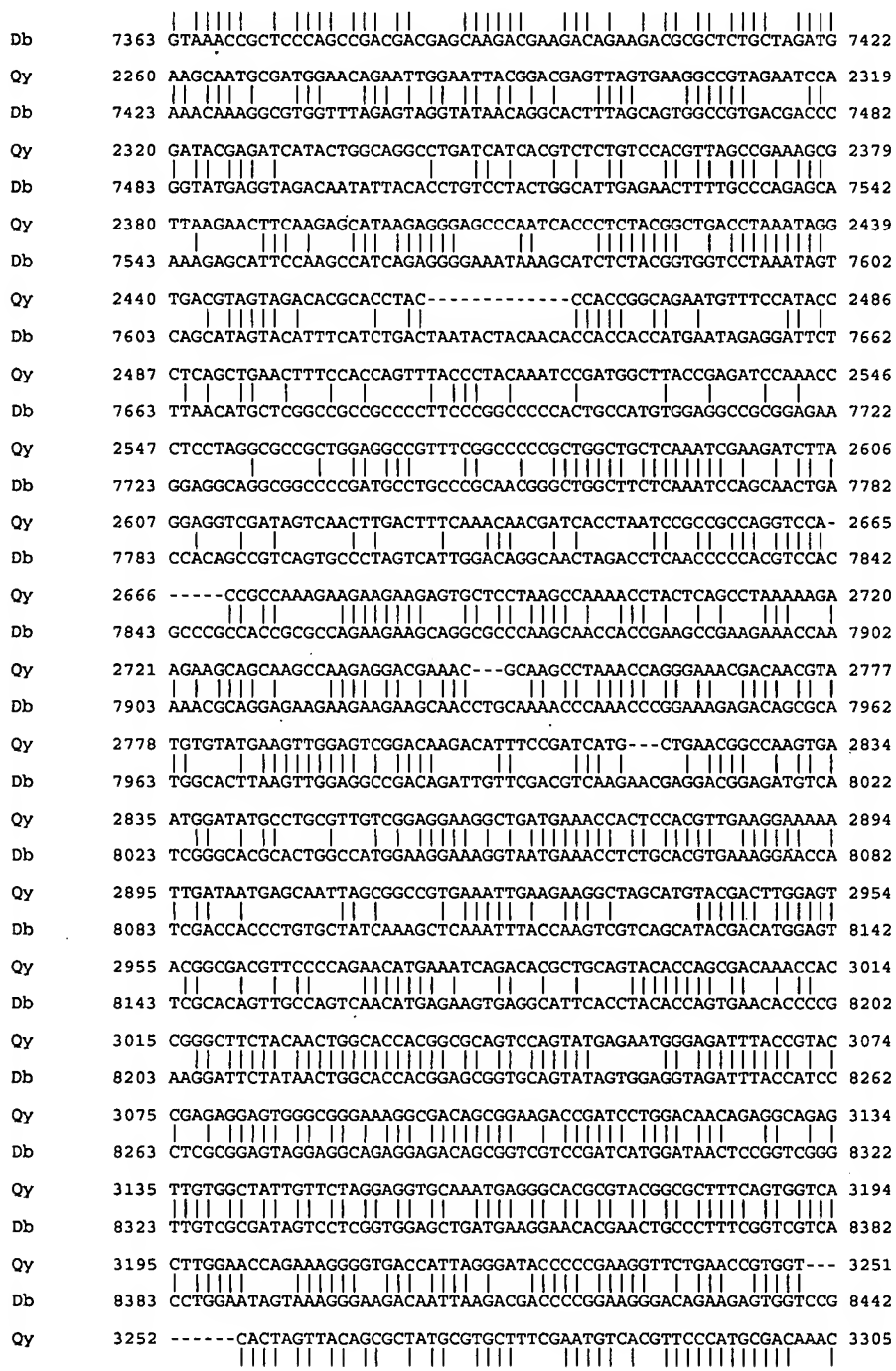
```
                     AVPSAMQNTLQNVLIAATKRNCNVTQMRELPTLDSATFNVECFRKYACNDEYWEEFAR
                     KPIRITTEFVTAYVARLKGPKAAALFAKTHNLVPLQEVPMDRFVMDMKRDVKVTPGTK
                     HTEERPKVQVIQAAEPLATAYLCGIHRELVRRLTAVLLPNIHTLFDMSAEDFDAIIAE
                     HFKQGDPVLETDIASFDKSQDDAMALTGLMILEDLGVDQPLLDLIECAFGEISSTHLP
                     TGTRFKFGAMMKSGMFLTLFVNTVLNVVIASRVLEERLKTSKCAAFIGDDNIIHGVVS
                     DKEMAERCATWLNMEVKIIDAVIGERPPYFCGGFILQDSVTSTACRVADPLKRLFKLG
                     KPLPADDEQDEDRRRALLDETKAWFRVGITDTLAVAVATRYEVDNITPVLLALRTFAQ
                     SKRAFQAIRGEIKHLYGGPK"
     mat_peptide     60. .1679
                     /product="nsP1"
     mat_peptide     1680. .4100
                     /product="nsP2"
     mat_peptide     4098. .5729
                     /product="nsP3"
     mat_peptide     5730. .7559
                     /product="nsP4"
     CDS             7608. .11345
                     /codon_start=1
                     /product="structural polyprotein"
                     /protein_id="AAA86136.1"
                     /db_xref="GI:1125071"
                     /translation="MNRGFFNMLGRRPFPAPTAMWRPRRRRQAAPMPARNGLASQIQQ
                     LTTAVSALVIGQATRPQTPRPRPPPRQKKQAPKQPPKPKKPKTQEKKKKQPAKPKPGK
                     RQRMALKLEADRLFDVKNEDGDVIGHALAMEGKVMKPLHVKGTIDHPVLSKLKFTKSS
                     AYDMEFAQLPVNMRSEAFTYTSEHPEGFYNWHHGAVQYSGGRFTIPRGVGGRGDSGRP
                     IMDNSGRVVAIVLGGADEGTRTALSVVTWNSKGKTIKTTPEGTEEWSAAPLVTAMCLL
                     GNVSFPCNRPPTCYTREPSRALDILEENVNHEAYDTLLNAILRCGSSGRSKRSVTDDF
                     TLTSPYLGTCSYCHHTEPCFSPIKIEQVWDEADDNTIRIQTSAQFGYDQSGAASSNKY
                     RYMSLEQDHTVKEGTMDDIKISTSGPCRRLSYKGYFLLAKCPPGDSVTVSIASSNSAT
                     SCTMARKIKPKFVGREKYDLPPVHGKKIPCTVYDRLKETTAGYITMHRPGPHAYTSYL
                     EESSGKVYAKPPSGKNITYECKCGDYKTGTVTTRTEITGCTAIKQCVAYKSDQTKWVF
                     NSPDSIRHADHTAQGKLHLPFKLIPSTCMVPVAHAPNVVHGFKHISLQLDTDHLTLLT
                     TRRLGANPEPTTEWIIGNTVRNFTVDRDGLEYIWGNHEPVRVYAQESAPGDPHGWPHE
                     IVQHYYHRHPVYTILAVASAAVAMMIGVTVAALCACKARRECLTPYALAPNAVIPTSL
                     ALLCCVRSANAETFTETMSYLWSNSQPFFWVQLCIPLAAVVVLMRCCSCCLPFLVVAG
                     AYLAKVDAYEHATTVPNVPQIPYKALVERAGYAPLNLEITVMSSEVLPSTNQEYITCK
                     FTTVVPSPKVRCCGSLECQPAAHADYTCKVFGGVYPFMWGGAQCFCDSENSQMSEAYV
                     ELSVDCATDHAQAIKVHTAAMKVGLRIVYGNTTSFLDVYVNGVTPGTSKDLKVIAGPI
                     SALFTPFDHKVVINRGLVYNYDFPEYGAMKPGAFGDIQATSLTSKDLIASTDIRLLKP
                     SAKNVHVPYTQAASGFEMWKNNSGRPLQETAPFGCKIAVNPLRAVDCSYGNIPISIDI
                     PNAAFIRTSDAPLVSTVKCDVSECTYSADFGGMATLQYVSDREGQCPVHSHSSTATLQ
                     ESTVHVLEKGAVTVHFSTASPQANFIVSLCGKKTTCNAECKPPADHIVSTPHKNDQEF
                     QAAISKTSWSWLFALFGGASSLLIIGLMIFACSMMLTSTRR"
     mat_peptide     7608. .8399
                     /product="capsid"
     mat_peptide     8400. .8591
                     /product="E3"
     mat_peptide     8592. .9860
                     /product="E2"
     mat_peptide     9861. .10025
                     /product="6K"
     mat_peptide     10026. .11342
                     /product="E1"
     3'UTR           11346. .11663
BASE COUNT     3291 a   3105 c   2907 g   2360 t
ORIGIN

  Query Match           36.9%;  Score 2396;  DB 14;  Length 11663;
  Best Local Similarity   64.6%;  Pred. No. 0;
  Matches 3751;  Conservative    0;  Mismatches 1985;  Indels   74;  Gaps    9;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
              |||| |     |  |  || | ||| || || ||  |||| |||||||  || ||| | |
Db       5707 ACTGTCTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5766

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
              || |||| || || |  ||      | ||   |||||| |  ||||| ||    || | |
Db       5767 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTTCTGG 5826

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
              | |     |||||||||   |||||        | |||| ||  | |  |     |
Db       5827 AAAGAATCTACGCCCCGGTGCTCGACACGTCGAAAGAGGAACAGCTCAAACTCAGGTACC 5886

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
              | |      || | |||| || | |||||||||| || || ||||||||||||||   ||
Db       5887 AGATGATGCCCACCGAAGCCAACAAAAGCAGGTACCAGTCTCGAAAAGTAGAAAACCAGA 5946

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
              |||| || ||  | |||||||  |||| ||  |  |    |||   ||  |  ||
Db       5947 AAGCCATAACCACTGAGCGACTGCTTTCAGGGCTACGACTGTATAACTCTGCCACAG--- 6003

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
              |||      || || || |   ||| || ||   ||||   || ||    |  |||
Db       6004 ATCAGCCAGAATGCTATAAGATCACCTACCCGAAACCATCGTATTCCAGCAGTGTACCAG 6063

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
              | |  |      || || | || ||       || |||     | | || |||||||
Db       6064 CGAACTACTCTGACCCAAAGTTTGCTGTAGCTGTTTGTAACAACTATCTGCATGAGAATT 6123

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
              |||| || |||||   ||||   || || || || |||||||| ||  | || ||||| |
Db       6124 ACCCGACGGTAGCATCTTATCAGATCACCGACGAGTACGATGCTTACTTGGATATGGTAG 6183

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
              ||||  ||      || |||||||| || |||||||| || || || || ||||| ||||
Db       6184 ACGGGACAGTCGCTTGCCTAGATACTGCAACTTTTTGCCCCGCCAAGCTTAGAAGTTACC 6243

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
              | || |  ||     |||      ||| | || |     || || ||||  | || ||||
Db       6244 CGAAAAGACACGAGTATAGAGCCCCAAACATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6303

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
              | || || ||||  ||  |   ||| || |||||||| || |||||||| || |||||||
Db       6304 ACACGTTGCAAAACGTGCTCATTGCCGCGACTAAAAGAAACTGCAACGTCACACAAATGC 6363

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
              | ||| | ||     | || || ||| |||| || ||||| || ||     ||||| ||||
Db       6364 GTGAACTGCCAACACTGGACTCAGCGACATTCAACGTTGAATGCTTTCGAAAATATGCAT 6423

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
              ||||||| ||||| |||||   ||| |  || || ||| || | ||||| ||| ||
Db       6424 GCAATGACGAGTATTGGGAGGAGTTTGCCCGAAAGCCAATTAGGATCACTACTGAGTTCG 6483

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
              ||||   ||| ||| | |  |||||||| || || || || ||| |||| || || || |
Db       6484 TTACCGCATACGTGGCCAGACTGAAAGGCCCTAAGGCCGCCGCACTGTTCGCAAAGACGC 6543

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
              ||||| |    || ||||| ||  | || ||||||  |||||||||||  | || ||||
Db       6544 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6603

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
              | || |||||||| || |||||||||| ||||| ||  | || || || || || || | |
Db       6604 ACGTGAAAGTTACACCTGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6663

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
              | || ||||| ||||| || |||||||||| |||||||||| ||||| ||||| || |
Db       6664 AAGCCGCAGAACCCCTGGCGACCGCTTACCTATGCGGGATCCACCGGGAGTTAGTGCGCA 6723

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
              | || |  || ||  | || || || || || || || ||||| |||||||| ||||| || |
Db       6724 GGCTTACAGCCGTTTGCTACCCAACATTCACACGCTCTTGACATGTCGGCGGAGGACT 6783

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
              ||||||| || || || |||| || | || || ||||| ||| |||| ||||| ||||
Db       6784 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGTGACCCGGTACTGGAGACGGATATCG 6843

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
              | ||||| || |||||| |||||||||||||| ||  |  | |  |||||||| | || |
Db       6844 CCTCGTTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGCCTGATGATCTTGGAAG 6903

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
              || | ||||| |||||||| ||  | || |||| |||   || || ||  | ||  |||
Db       6904 ACCTGGGTGTGGACCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 6963

Qy       1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
```

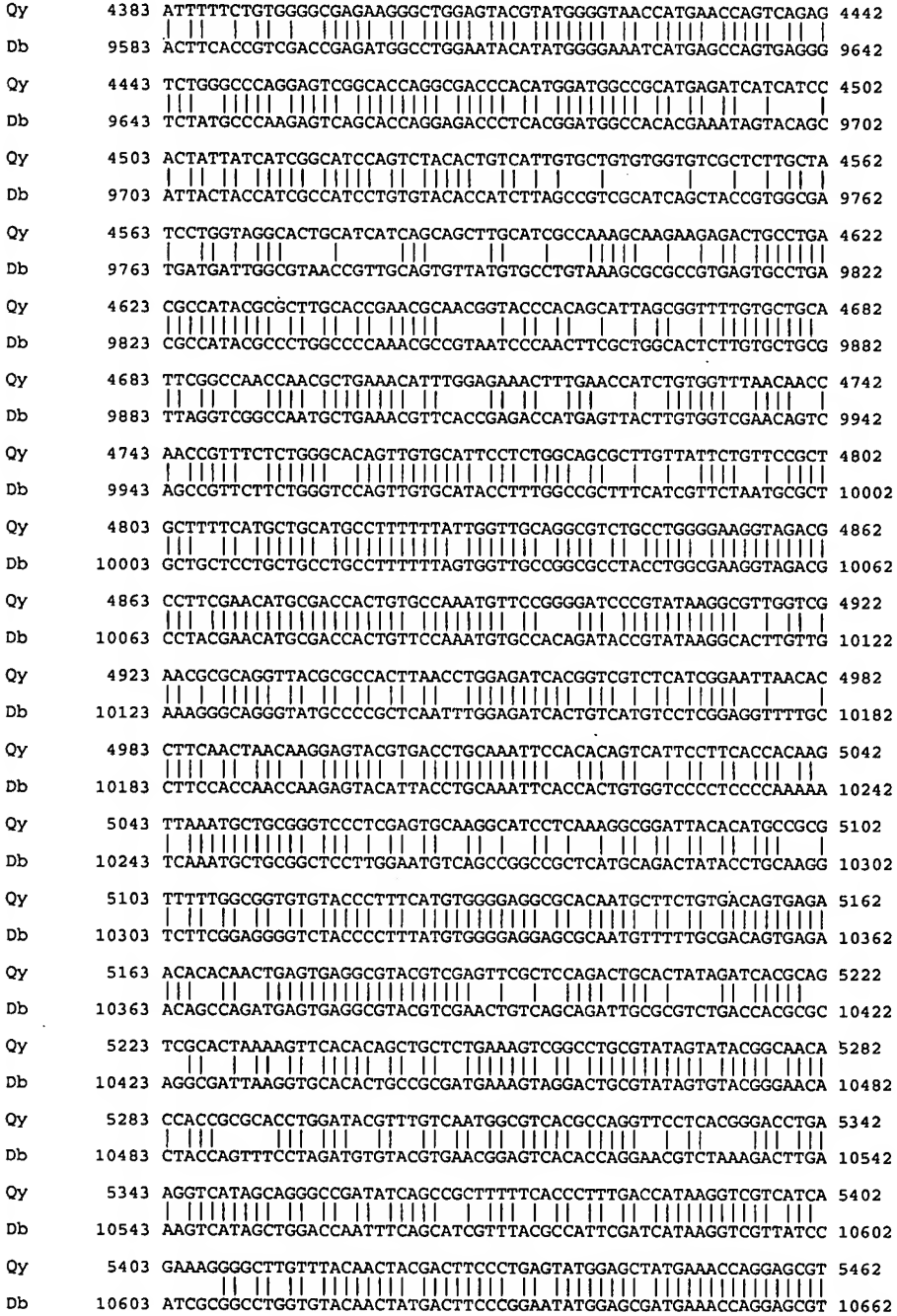





Qy 4023 CCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGATGA 4082
Db 9184 CGTACGAGTGCAAGTGCGGCGATTACAAGACCGGAACCGTTACGACCCGTACCGAAATCA 9243
Qy 4083 ACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGTCT 4142
Db 9244 CGGGCTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAACGAAGTGGGTCT 9303
Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9304 TCAACTCGCCGGACTCGATCAGACACGCCGACCACACGGCCCAAGGGAAATTGCATTTGC 9363
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9364 CTTTCAAGCTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAACGTAGTAC 9423
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9424 ACGGCTTTAAACACATCAGCCTCCAATTAGACACAGACCATCTGACATTGCTCACCACCA 9483
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9484 GGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCATCGGAAACACGGTTAGAA 9543
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9544 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGCAATCACGAACCAGTAAGGG 9603
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db 9604 TCTATGCCCAAGAGTCTGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9663
Qy 4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db 9664 ATTACTATCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTGCTGTGGCGA 9723
Qy 4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db 9724 TGATGATTGGCGTAACTGTTGCAGCATTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9783
Qy 4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db 9784 CGCCATATGCCCTGGCCCCAAATGCCGTGATTCCAACTTCGCTGGCACTTTTGTGCTGTG 9843
Qy 4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db 9844 TTAGGTCGGCTAATGCTGAAACATTCACCGAGACCATGAGTTACTTATGGTCGAACAGCC 9903
Qy 4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
Db 9904 AGCCGTTCTTCTGGGTCCAGCTGTGTATACCTCTGGCCGCTGTCGTCGTTCTAATGCGCT 9963
Qy 4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
Db 9964 GTTGCTCATGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10023
Qy 4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
Db 10024 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10083
Qy 4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db 10084 AAAGGGCAGGGTACGCCCCGCTCAATTTGGAGATTACTGTCATGTCCTCGGAGGTTTTGC 10143
Qy 4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db 10144 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCTAAAG 10203
Qy 5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db 10204 TCAGATGCTGCGGCTCCTTGGAATGTCAGCCCGCCGCTCACGCAGACTATACCTGCAAGG 10263
Qy 5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db 10264 TCTTTGGAGGGGTGTACCCCTTCATGTGGGGAGGAGCACAATGTTTTTGCGACAGTGAGA 10323
Qy 5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db 10324 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGTAGATTGCGCGACTGACCACGCGC 10383
Qy 5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db 10384 AGGCGATTAAGGTGCATACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10443
Qy 5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db 10444 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACCTGA 10503
Qy 5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db 10504 AAGTCATAGCTGGACCAATTTCAGCATTGTTTACACCATTCGATCACAAGGTCGTTATCA 10563
Qy 5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db 10564 ATCGCGGCCTGGTGTACAACTATGACTTTCCGGAATACGGAGCGATGAAACCAGGAGCGT 10623
Qy 5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db 10624 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAAGACCTCATCGCCAGCACAGACATTA 10683
Qy 5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db 10684 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCTGGAT 10743
Qy 5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db 10744 TCGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCCCCTTTTGGGTGCA 10803
Qy 5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db 10804 AGATTGCAGTCAATCCGCTTCGAGCGGTGGACTGCTCATACGGGAACATTCCCATTTCTA 10863
Qy 5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db 10864 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10923
Qy 5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db 10924 AATGTGATGTCAGTGAGTGCACTTATTCAGCGGACTTCGGAGGGATGGCTACCCTGCAGT 10983
Qy 5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db 10984 ATGTATCCGACCGCGAAGGACAATGCCCTGTACATTCGCATTCGAGCACAGCAACCCTCC 11043
Qy 5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db 11044 AAGAGTCGACAGTTCATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTCAGCACCGCGA 11103
Qy 5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db 11104 GCCCACAGGCGAACTTCATTGTATCGCTGTGTGGTAAGAAGACAACATGCAATGCAGAAT 11163
Qy 6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db 11164 GCAAACCACCAGCTGATCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTCCAAG 11223
Qy 6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db 11224 CCGCCATCTCAAAAACTTCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11283
Qy 6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
Db 11284 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11343
Qy 6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241

```
Db      11344 GACCGCTACGCCCCAATGACCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11403

Qy       6242 GCATAATGCCACGCGCCGCT----------------------------------TGACACT 6268

Db      11404 GCATAATGCATCAGGCTGGTATATTAGATCCCCGCTTACCGCGGGCAATATAGCAACACC 11463

Qy       6269 AAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGT 6318

Db      11464 AAAACTCGACGTATTTCCGAGGAAGCGCAGTGCATAATGCTGCGCAGTGT 11513


RESULT 13
AR042412
LOCUS       AR042412               11703 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 8 from patent US 5811407.
ACCESSION   AR042412
VERSION     AR042412.1  GI:5962908
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 11703)
  AUTHORS   Johnston,R.E., Davis,N.L. and Simpson,D.A.
  TITLE     System for the in vivo delivery and expression of heterologous
            genes in the bone marrow
  JOURNAL   Patent: US 5811407-A 8 22-SEP-1998;
FEATURES             Location/Qualifiers
     source          1. .11703
                     /organism="unknown"
BASE COUNT     3306 a   3051 c   2913 g   2433 t
ORIGIN

  Query Match           36.8%;  Score 2388.4;  DB 6;  Length 11703;
  Best Local Similarity   64.7%;  Pred. No. 0;
  Matches 3698;  Conservative    0;  Mismatches 1981;  Indels   41;  Gaps    8;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639

Db       5746 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5805

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699

Db       5806 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTCCTGG 5865

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759

Db       5866 AAAGAATTCATGCCCCGGTGCTCGACACGTCGAAAGAGGAACAACTCAAACTCAGGTACC 5925

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819

Db       5926 AGATGATGCCCACCGAAGCCAACAAAAGTAGGTACCAGTCTCGTAAAGTAGAAAATCAGA 5985

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879

Db       5986 AAGCCATAACCACTGAGCGACTACTGTCAGGACTACGACTGTATAACTCTGCCACAG--- 6042

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939

Db       6043 ATCAGCCAGAATGCTATAAGATCACCTATCCGAAACCATTGTACTCCAGTAGCGTACCGG 6102

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999

Db       6103 CGAACTACTCCGATCCACAGTTCGCTGTAGCTGTCTGTAACAACTATCTGCATGAGAACT 6162

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059

Db       6163 ATCCGACAGTAGCATCTTATCAGATTACTGACGAGTACGATGCTTACTTGGATATGGTAG 6222

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119

Db       6223 ACGGGACAGTCGCCTGCCTGGATACTGCAACCTTCTGCCCCGCTAAGCTTAGAAGTTACC 6282

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179

Db       6283 CGAAAAAACATGAGTATAGAGCCCCGAATATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6342

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239

Db       6343 ACACGCTACAAAATGTGCTCATTGCCGCAACTAAAAGAAATTGCAACGTCACGCAGATGC 6402

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299

Db       6403 GTGAACTGCCAACACTGGACTCAGCGACATTCAATGTCGAATGCTTTCGAAAATATGCAT 6462

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359

Db       6463 GTAATGACGAGTATTGGGAGGAGTTCGCTCGGAAGCCAATTAGGATTACCACTGAGTTTG 6522

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419

Db       6523 TCACCGCATATGTAGCTAGACTGAAAGGCCCTAAGGCCGCCGCACTATTTGCAAAGACGT 6582

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479

Db       6583 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6642

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539

Db       6643 ACGTGAAAGTTACACCAGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6702

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599

Db       6703 AAGCCGCAGAACCCCTGGCGACTGCTTACTTATGCGGGATTCACCGGGAATTAGTGCGTA 6762

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659

Db       6763 GGCTTACGGCCGTCTTGCTTCCAAACATTCACACGCTTTTTGACATGTCGGCGGAGGATT 6822

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719

Db       6823 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGCGACCCGGTACTGGAGACGGATATCG 6882

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779

Db       6883 CATCATTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGTCTGATGATCTTGGAGG 6942

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839

Db       6943 ACCTGGGTGTGGATCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 7002

Qy       1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899

Db       7003 CCACCCATCTACCTACGGGTACTCGTTTTAAATTCGGGGCGATGATGAAATCCGGAATGT 7062

Qy       1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959

Db       7063 TCCTCACACTTTTTGTCAACACAGTTTTGAATGTCGTTATCGCCAGCAGAGTACTAGAAG 7122

Qy       1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019

Db       7123 AGCGGCTTAAAACGTCCAGATGTGCAGCGTTCATTGGCGACGACAACATCATACATGGAG 7182

Qy       2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079

Db       7183 TAGTATCTGACAAAGAAATGGCTGAGAGGTGCGCCACCTGGCTCAACATGGAGGTTAAGA 7242

Qy       2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139

Db       7243 TCATCGACGCAGTCATCGGTGAGAGACCACCTTACTTCTGCGGCGGATTTATCTTGCAAG 7302

Qy       2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199

Db       7303 ATTCGGTTACTTCCACAGCGTGCCGCGTGGCGGACCCCCTGAAAAGGCTGTTTAAGTTGG 7362

Qy       2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
```

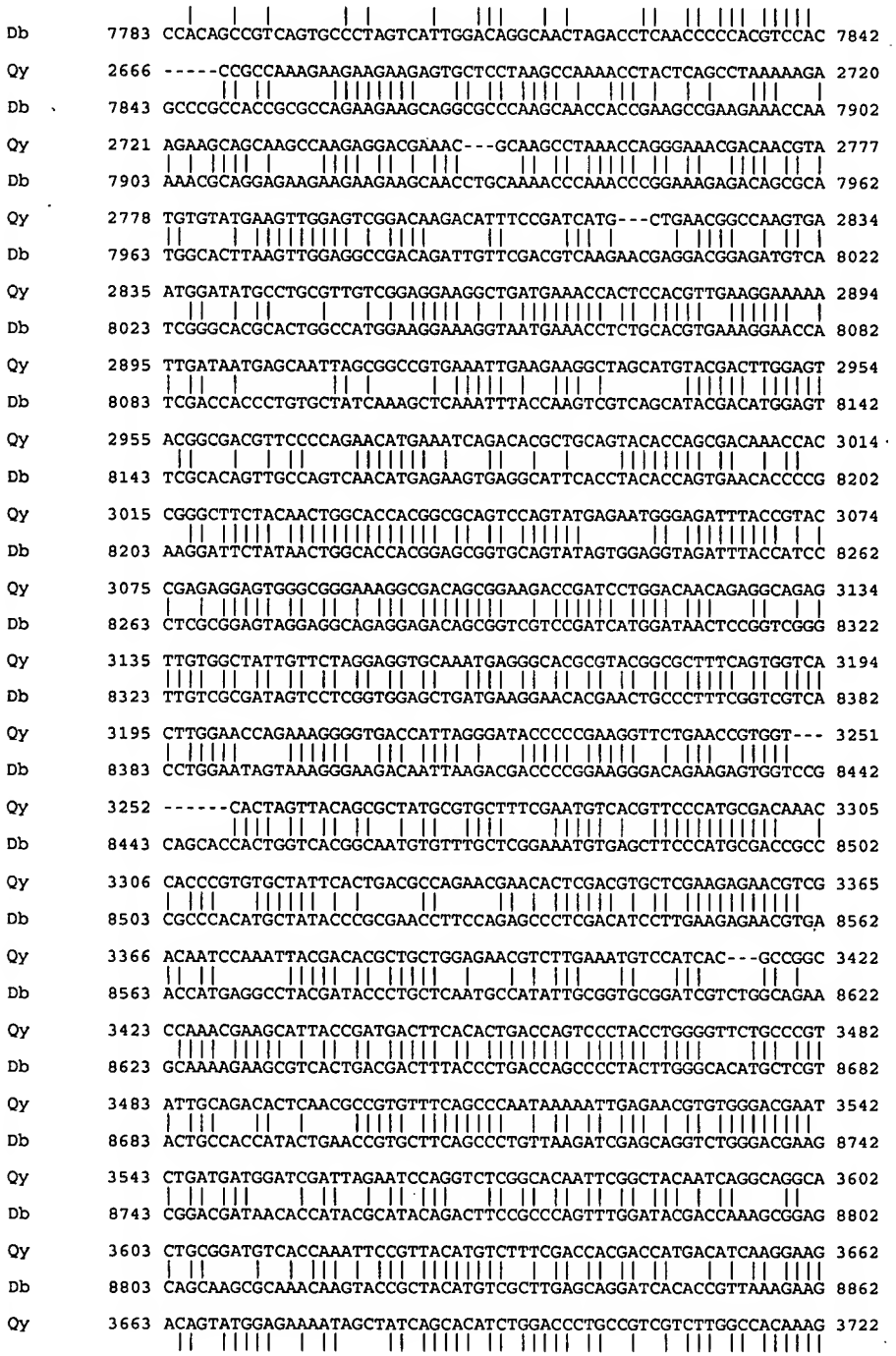

Db 7363 GTAAACCGCTCCCAGCCGACGACGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 7422
Qy 2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319
Db 7423 AAACAAAGGCGTGGTTTAGAGTAGGTATAACAGGCACTTTAGCAGTGGCCGTGACGACCC 7482
Qy 2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379
Db 7483 GGTATGAGGTAGACAATATTACACCTGTCCTACTGGCATTGAGAACTTTTGCCCAGAGCA 7542
Qy 2380 TTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGG 2439
Db 7543 AAAGAGCATTCCAAGCCATCAGAGGGGAAATAAAGCATCTCTACGGTGGTCCTAAATAGT 7602
Qy 2440 TGACGTAGTAGACACGCACCTAC----------CCACCGGCAGAATGTTTCCATACC 2486
Db 7603 CAGCATAGTACATTTCATCTGACTAATACTACAACACCACCACCATGAATAGAGGATTCT 7662
Qy 2487 CTCAGCTGAACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACC 2546
Db 7663 TTAACATGCTCGGCCGCCGCCCCTTCCCGGCCCCCACTGCCATGTGGAGGCCGCGGAGAA 7722
Qy 2547 CTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTA 2606
Db 7723 GGAGGCAGGCGGCCCCGATGCCTGCCCGCAACGGGCTGGCTTCTCAAATCCAGCAACTGA 7782
Qy 2607 GGAGGTCGATAGTCAACTTGACTTTCAAACAACGATCACCTAATCCGCCGCCAGGTCCA- 2665
Db 7783 CCACAGCCGTCAGTGCCCTAGTCATTGGACAGGCAACTAGACCTCAACCCCCACGTCCAC 7842
Qy 2666 -----CCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGA 2720
Db 7843 GCCCGCCACCGCGCCAGAAGAAGCAGGCGCCCAAGCAACCACCGAAGCCGAAGAAACCAA 7902
Qy 2721 AGAAGCAGCAAGCCAAGAGGACGAAAC---GCAAGCCTAAACCAGGGAAACGACAACGTA 2777
Db 7903 AAACGCAGGAGAAGAAGAAGAAGCAACCTGCAAAACCCAAACCCGGAAAGAGACAGCGCA 7962
Qy 2778 TGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATCATG---CTGAACGGCCAAGTGA 2834
Db 7963 TGGCACTTAAGTTGGAGGCCGACAGATTGTTCGACGTCAAGAACGAGGACGGAGATGTCA 8022
Qy 2835 ATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACTCCACGTTGAAGGAAAAA 2894
Db 8023 TCGGGCACGCACTGGCCATGGAAGGAAAGGTAATGAAACCTCTGCACGTGAAAGGAACCA 8082
Qy 2895 TTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAGCATGTACGACTTGGAGT 2954
Db 8083 TCGACCACCCTGTGCTATCAAAGCTCAAATTTACCAAGTCGTCAGCATACGACATGGAGT 8142
Qy 2955 ACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTACACCAGCGACAAACCAC 3014
Db 8143 TCGCACAGTTGCCAGTCAACATGAGAAGTGAGGCATTCACCTACACCAGTGAACACCCCG 8202
Qy 3015 CGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAATGGGAGATTTACCGTAC 3074
Db 8203 AAGGATTCTATAACTGGCACCACGGAGCGGTGCAGTATAGTGGAGGTAGATTTACCATCC 8262
Qy 3075 CGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGACAACAGAGGCAGAG 3134
Db 8263 CTCGCGGAGTAGGAGGCAGAGGAGACAGCGGTCGTCCGATCATGGATAACTCCGGTCGGG 8322
Qy 3135 TTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGCGCTTTCAGTGGTCA 3194
Db 8323 TTGTCGCGATAGTCCTCGGTGGAGCTGATGAAGGAACACGAACTGCCCTTTCGGTCGTCA 8382
Qy 3195 CTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTCTGAACCGTGGT--- 3251
Db 8383 CCTGGAATAGTAAAGGGAAGACAATTAAGACGACCCCGGAAGGGACAGAAGAGTGGTCCG 8442
Qy 3252 ------CACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGACAAAC 3305
Db 8443 CAGCACCACTGGTCACGGCAATGTGTTTGCTCGGAAATGTGAGCTTCCCATGCGACCGCC 8502
Qy 3306 CACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAACGTCG 3365
Db 8503 CGCCCACATGCTATACCCGCGAACCTTCCAGAGCCCTCGACATCCTTGAAGAGAACGTGA 8562
Qy 3366 ACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAATGTCCATCAC---GCCGGC 3422
Db 8563 ACCATGAGGCCTACGATACCCTGCTCAATGCCATATTGCGGTGCGGATCGTCTGGCAGAA 8622
Qy 3423 CCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCCCGT 3482
Db 8623 GCAAAAGAAGCGTCACTGACGACTTTACCCTGACCAGCCCCTACTTGGGCACATGCTCGT 8682
Qy 3483 ATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACGAAT 3542
Db 8683 ACTGCCACCATACTGAACCGTGCTTCAGCCCTGTTAAGATCGAGCAGGTCTGGGACGAAG 8742
Qy 3543 CTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAGGCA 3602
Db 8743 CGGACGATAACACCATACGCATACAGACTTCCGCCCAGTTTGGATACGACCAAAGCGGAG 8802
Qy 3603 CTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAAGGAAG 3662
Db 8803 CAGCAAGCGCAAACAAGTACCGCTACATGTCGCTTGAGCAGGATCACACCGTTAAAGAAG 8862
Qy 3663 ACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCACAAAG 3722
Db 8863 GCACCATGGATGACATCAAGATTAGCACCTCAGGACCGTGTAGAAGGCTTAGCTACAAAG 8922
Qy 3723 GGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCACGAGCG 3782
Db 8923 GATACTTTCTCCTCGCAAAATGCCCTCCAGGGGACAGCGTAACGGTTAGCATAGTGAGTA 8982
Qy 3783 GAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTAGAG 3842
Db 8983 GCAACTCAGCAACGTCATGTACACTGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGG 9042
Qy 3843 AGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATCACT 3902
Db 9043 AAAAATATGATCTACCTCCCGTTCACGGTAAAAAAATTCCTTGCACAGTGTACGACCGTC 9102
Qy 3903 TGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAAGT 3962
Db 9103 TGAAAGAAACAACTGCAGGCTACATCACTATGCACAGGCCGGGACCGCACGCTTATACAT 9162
Qy 3963 CCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGTCA 4022
Db 9163 CCTACCTGGAAGAATCATCAGGGAAAGTTTACGCAAAGCCGCCATCTGGGAAGAACATTA 9222
Qy 4023 CCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGATGA 4082
Db 9223 CGTATGAGTGCAAGTGCGGCGACTACAAGACCGGAACCGTTTCGACCCGCACCGAAATCA 9282
Qy 4083 ACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGTCT 4142
Db 9283 CTGGTTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAACGAAGTGGGTCT 9342
Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9343 TCAACTCACCGGACTTGATCAGACATGACGACCACACGGCCCAAGGGAAATTGCATTTGC 9402
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9403 CTTTCAAGTTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAATGTAATAC 9462
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9463 ATGGCTTTAAACACATCAGCCTCCAATTAGATACAGACCACTTGACATTGCTCACCACCA 9522
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9523 GGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCGTCGGAAAGACGGTCAGAA 9582

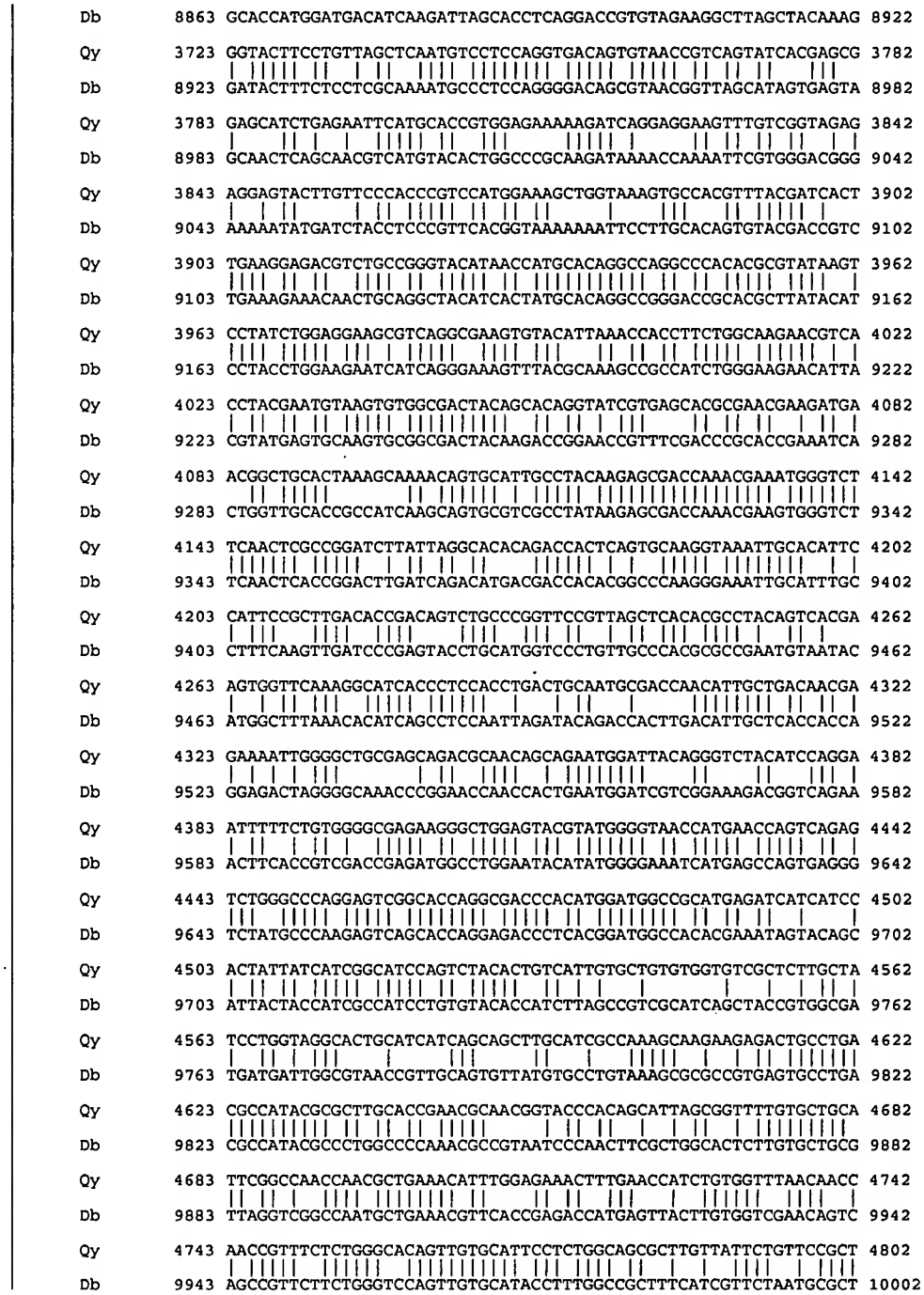


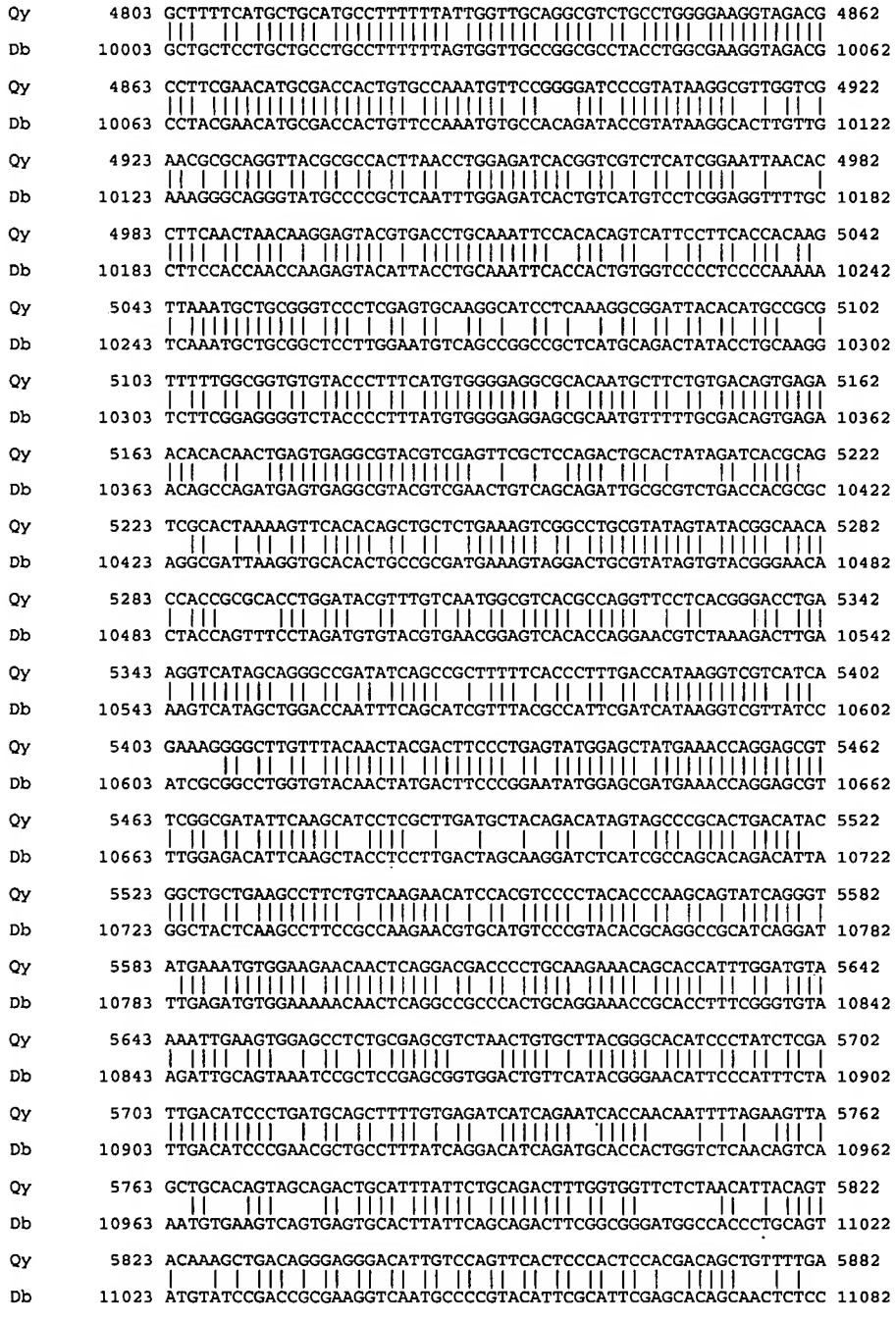
RESULT 14
AR096558
LOCUS AR096558 11703 bp DNA linear PAT 08-SEP-2000
DEFINITION Sequence 8 from patent US 6008035.
ACCESSION AR096558
VERSION AR096558.1 GI:10025468
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 11703)
AUTHORS Johnston,R.E., Davis,N.L. and Simpson,D.A.
TITLE System for the in vivo delivery and expression of heterologous genes in the bone marrow
JOURNAL Patent: US 6008035-A 8 28-DEC-1999;
FEATURES Location/Qualifiers

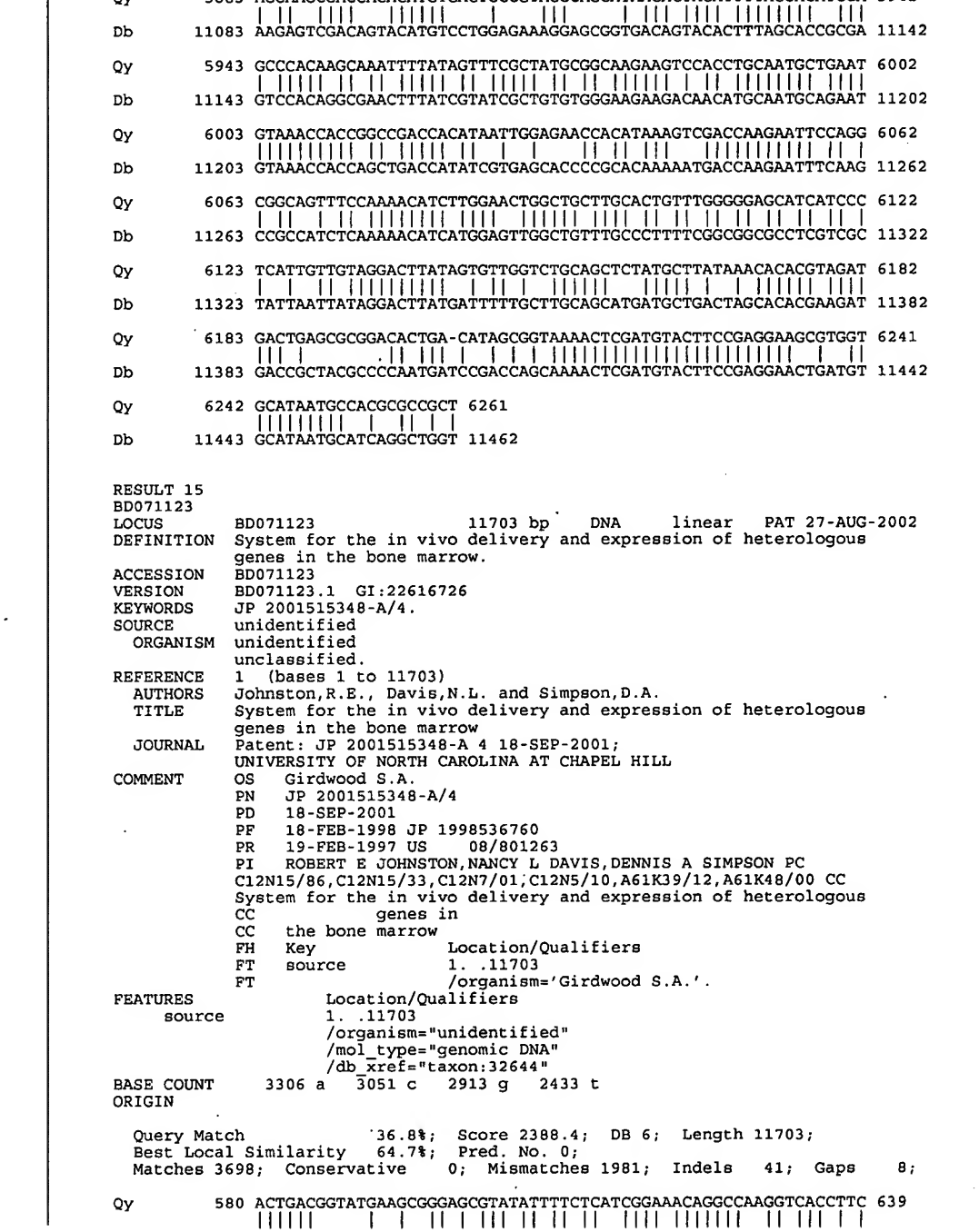
```
     source          1. .11703
                     /organism="unknown"
BASE COUNT     3306 a    3051 c    2913 g    2433 t
ORIGIN

  Query Match             36.8%;  Score 2388.4;  DB 6;  Length 11703;
  Best Local Similarity   64.7%;  Pred. No. 0;
  Matches 3698;  Conservative    0;  Mismatches 1981;  Indels   41;  Gaps    8;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
              ||||||      |  |  || | ||| || || ||  |||| |||||||  || ||| | |
Db       5746 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5805

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
              || |||| || || |  ||      | ||   ||||||  |  |||||  ||     ||||  |
Db       5806 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTCCTGG 5865

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
              | |    | | ||||||   |||||      | ||||  ||    |  |    |     |
Db       5866 AAAGAATTCATGCCCCGGTGCTCGACACGTCGAAAGAGGAACAACTCAAACTCAGGTACC 5925

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
              |  |      || | ||||  || | ||| ||||| || || ||  ||||||||||||  ||
Db       5926 AGATGATGCCCACCGAAGCCAACAAAAGTAGGTACCAGTCTCGTAAAGTAGAAAATCAGA 5985

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
              |||| || ||  | |||||||||  | || ||| |  |     |||   ||  |   ||
Db       5986 AAGCCATAACCACTGAGCGACTACTGTCAGGACTACGACTGTATAACTCTGCCACAG--- 6042

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
              |||      || || || |   |||  |||||   |||| | |||||     |  |||
Db       6043 ATCAGCCAGAATGCTATAAGATCACCTATCCGAAACCATTGTACTCCAGTAGCGTACCGG 6102

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
                |   |    |  | || || |||| ||        || |||    | | || ||||| |
Db       6103 CGAACTACTCCGATCCACAGTTCGCTGTAGCTGTCTGTAACAACTATCTGCATGAGAACT 6162

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
              | || ||||||||   ||||   || || || || |||||||| ||  | || |||||  |
Db       6163 ATCCGACAGTAGCATCTTATCAGATTACTGACGAGTACGATGCTTACTTGGATATGGTAG 6222

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
              ||||  ||     ||| || ||||| || || || || || ||||| || ||||| ||||
Db       6223 ACGGGACAGTCGCCTGCCTGGATACTGCAACCTTCTGCCCCGCTAAGCTTAGAAGTTACC 6282

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
              | || || |||    |||        ||  | ||  |    || || |||||  | || ||||
Db       6283 CGAAAAAACATGAGTATAGAGCCCCGAATATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6342

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
              | ||  ||||||||||   |   ||| || |||||||| ||||||||||| || || ||||
Db       6343 ACACGCTACAAAATGTGCTCATTGCCGCAACTAAAAGAAATTGCAACGTCACGCAGATGC 6402

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
              | ||| | ||     | || || ||| |||| ||||| || || ||    ||||| ||||
Db       6403 GTGAACTGCCAACACTGGACTCAGCGACATTCAATGTCGAATGCTTTCGAAAATATGCAT 6462

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
              | ||||| ||||| |||||    ||        || || ||| || | || || |||    |
Db       6463 GTAATGACGAGTATTGGGAGGAGTTCGCTCGGAAGCCAATTAGGATTACCACTGAGTTTG 6522

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
              | ||   ||||||  | |  |||||||| || || || || ||| | ||||| || ||
Db       6523 TCACCGCATATGTAGCTAGACTGAAAGGCCCTAAGGCCGCCGCACTATTTGCAAAGACGT 6582

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
              ||||| |    || |||||  ||  | || ||||||  ||||||||||||  | || ||||
Db       6583 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6642

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
              | || |||||||| || ||||||||||| ||||| |  | || || || || || || |
Db       6643 ACGTGAAAGTTACACCAGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6702
```

```
Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
              | || ||||| ||||| || || |||||| | |||||||| || ||||||||||| ||||
Db       6703 AAGCCGCAGAACCCCTGGCGACTGCTTACTTATGCGGGATTCACCGGGAATTAGTGCGTA 6762

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
              | || |  || ||  | || ||||| || || || || || |||||||| |||| ||||
Db       6763 GGCTTACGGCCGTCTTGCTTCCAAACATTCACACGCTTTTTGACATGTCGGCGGAGGATT 6822

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
              ||||||| || || || ||||| ||| | || |||||||| ||| |||| ||||| |||| .
Db       6823 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGCGACCCGGTACTGGAGACGGATATCG 6882

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
              | || || || |||||| |||||||||||||| ||  |  | |   |||||||| | ||||
Db       6883 CATCATTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGTCTGATGATCTTGGAGG 6942

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
              || | ||||| || ||||| ||  | || |||| |||   || || ||  | ||  |||
Db       6943 ACCTGGGTGTGGATCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 7002

Qy       1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
              |   || |||||||| || ||  | |||||||| || || ||||||||||||| ||||
Db       7003 CCACCCATCTACCTACGGGTACTCGTTTTAAATTCGGGGCGATGATGAAATCCGGAATGT 7062

Qy       1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
              || | || || |||||||||||| |  ||| || | || || ||||||||||||  |
Db       7063 TCCTCACACTTTTTGTCAACACAGTTTTGAATGTCGTTATCGCCAGCAGAGTACTAGAAG 7122

Qy       1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
              | ||| | |  |||||     || || || |  || |||||||| |||||  | ||||| |
Db       7123 AGCGGCTTAAAACGTCCAGATGTGCAGCGTTCATTGGCGACGACAACATCATACATGGAG 7182

Qy       2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
              | || || ||||     ||||| ||||| |||||||| ||||| |||||||| || || |
Db       7183 TAGTATCTGACAAAGAAATGGCTGAGAGGTGCGCCACCTGGCTCAACATGGAGGTTAAGA 7242

Qy       2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139
              | || || ||||| || |||   | | |||| |||||||| || ||||||||| ||    |
Db       7243 TCATCGACGCAGTCATCGGTGAGAGACCACCTTACTTCTGCGGCGGATTTATCTTGCAAG 7302

Qy       2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199
              |   | | ||   |||||| ||| | || || ||||| || |||||||| ||||||  | |
Db       7303 ATTCGGTTACTTCCACAGCGTGCCGCGTGGCGGACCCCCTGAAAAGGCTGTTTAAGTTGG 7362

Qy       2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
              | |||||  | |||| ||| ||     ||||||   ||| |  | || || ||||  |||
Db       7363 GTAAACCGCTCCCAGCCGACGACGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 7422

Qy       2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319
              || ||| |   |||  || || || || | | |||| |||||  ||
Db       7423 AAACAAAGGCGTGGTTTAGAGTAGGTATAACAGGCACTTTAGCAGTGGCCGTGACGACCC 7482

Qy       2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379
              | || |||  |                |  || |  | |  ||   || || ||| | |||
Db       7483 GGTATGAGGTAGACAATATTACACCTGTCCTACTGGCATTGAGAACTTTTGCCCAGAGCA 7542

Qy       2380 TTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGG 2439
                |   ||| |   ||| ||||||     ||     |||||||  | |||||||||
Db       7543 AAAGAGCATTCCAAGCCATCAGAGGGGAAATAAAGCATCTCTACGGTGGTCCTAAATAGT 7602

Qy       2440 TGACGTAGTAGACACGCACCTAC----------CCACCGGCAGAATGTTTCCATACC 2486
                | |||||  |      |  ||                |||||  ||  |     || |
Db       7603 CAGCATAGTACATTTCATCTGACTAATACTACAACACCACCACCATGAATAGAGGATTCT 7662

Qy       2487 CTCAGCTGAACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACC 2546
              | |  ||  |   |  |    | |||  |    |       |   |   | |
Db       7663 TTAACATGCTCGGCCGCCGCCCCTTCCCGGCCCCCACTGCCATGTGGAGGCCGCGGAGAA 7722

Qy       2547 CTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTA 2606
                    |    |  || |||    ||   |  |||||| ||||||||  |  | || |
Db       7723 GGAGGCAGGCGGCCCCGATGCCTGCCCGCAACGGGCTGGCTTCTCAAATCCAGCAACTGA 7782

Qy       2607 GGAGGTCGATAGTCAACTTGACTTTCAAACAACGATCACCTAATCCGCCGCCAGGTCCA- 2665
```

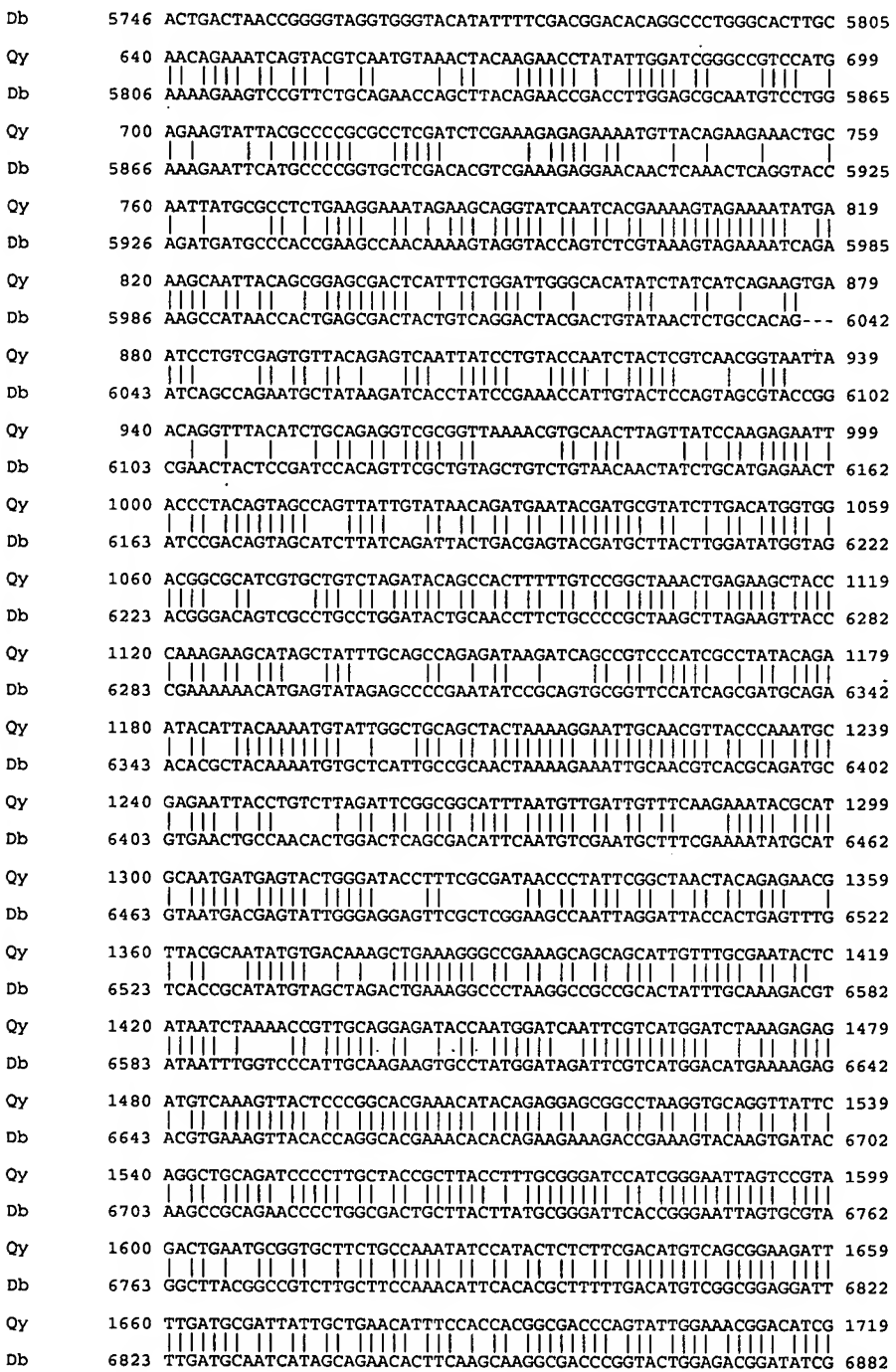


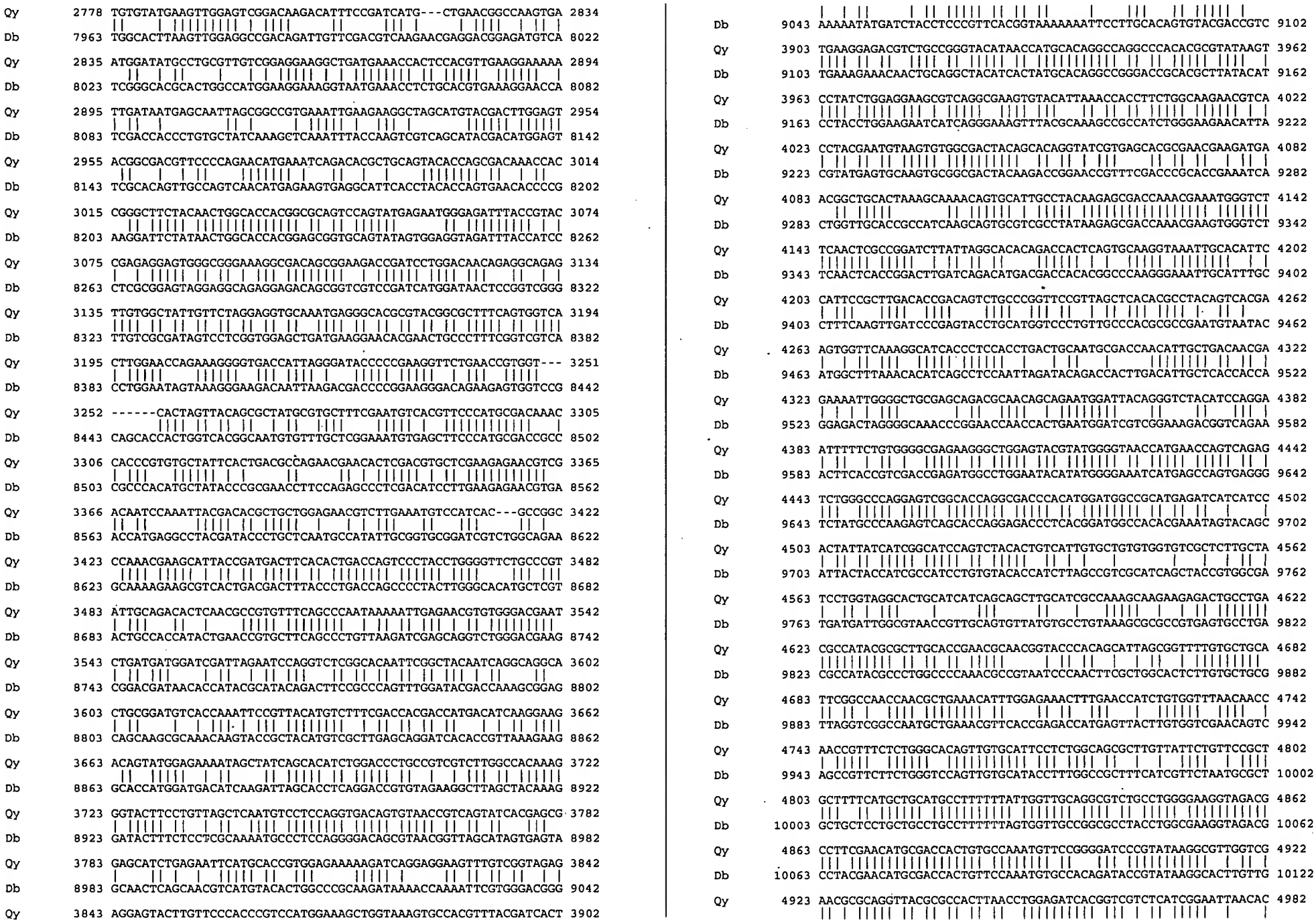






```
RESULT 15
BD071123
LOCUS       BD071123               11703 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  System for the in vivo delivery and expression of heterologous
            genes in the bone marrow.
ACCESSION   BD071123
VERSION     BD071123.1  GI:22616726
KEYWORDS    JP 2001515348-A/4.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 11703)
  AUTHORS   Johnston,R.E., Davis,N.L. and Simpson,D.A.
  TITLE     System for the in vivo delivery and expression of heterologous
            genes in the bone marrow
  JOURNAL   Patent: JP 2001515348-A 4 18-SEP-2001;
            UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
COMMENT     OS   Girdwood S.A.
            PN   JP 2001515348-A/4
            PD   18-SEP-2001
            PF   18-FEB-1998 JP 1998536760
            PR   19-FEB-1997 US    08/801263
            PI   ROBERT E JOHNSTON,NANCY L DAVIS,DENNIS A SIMPSON PC
            C12N15/86,C12N15/33,C12N7/01;C12N5/10,A61K39/12,A61K48/00 CC
            System for the in vivo delivery and expression of heterologous
            CC          genes in
            CC   the bone marrow
            FH   Key             Location/Qualifiers
            FT   source          1. .11703
            FT                   /organism='Girdwood S.A.'.
FEATURES             Location/Qualifiers
     source          1. .11703
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT     3306 a   3051 c   2913 g   2433 t
ORIGIN

  Query Match           36.8%;  Score 2388.4;  DB 6;  Length 11703;
  Best Local Similarity   64.7%;  Pred. No. 0;
  Matches 3698;  Conservative    0;  Mismatches 1981;  Indels   41;  Gaps    8;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
```


Db 5746 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5805
Qy 640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
Db 5806 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTCCTGG 5865
Qy 700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
Db 5866 AAAGAATTCATGCCCCGGTGCTCGACACGTCGAAAGAGGAACAACTCAAACTCAGGTACC 5925
Qy 760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
Db 5926 AGATGATGCCCACCGAAGCCAACAAAAGTAGGTACCAGTCTCGTAAAGTAGAAAATCAGA 5985
Qy 820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
Db 5986 AAGCCATAACCACTGAGCGACTACTGTCAGGACTACGACTGTATAACTCTGCCACAG--- 6042
Qy 880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
Db 6043 ATCAGCCAGAATGCTATAAGATCACCTATCCGAAACCATTGTACTCCAGTAGCGTACCGG 6102
Qy 940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
Db 6103 CGAACTACTCCGATCCACAGTTCGCTGTAGCTGTCTGTAACAACTATCTGCATGAGAACT 6162
Qy 1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
Db 6163 ATCCGACAGTAGCATCTTATCAGATTACTGACGAGTACGATGCTTACTTGGATATGGTAG 6222
Qy 1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
Db 6223 ACGGGACAGTCGCCTGCCTGGATACTGCAACCTTCTGCCCCGCTAAGCTTAGAAGTTACC 6282
Qy 1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
Db 6283 CGAAAAAACATGAGTATAGAGCCCCGAATATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6342
Qy 1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
Db 6343 ACACGCTACAAAATGTGCTCATTGCCGCAACTAAAAGAAATTGCAACGTCACGCAGATGC 6402
Qy 1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
Db 6403 GTGAACTGCCAACACTGGACTCAGCGACATTCAATGTCGAATGCTTTCGAAAATATGCAT 6462
Qy 1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
Db 6463 GTAATGACGAGTATTGGGAGGAGTTCGCTCGGAAGCCAATTAGGATTACCACTGAGTTTG 6522
Qy 1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
Db 6523 TCACCGCATATGTAGCTAGACTGAAAGGCCCTAAGGCCGCCGCACTATTTGCAAAGACGT 6582
Qy 1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
Db 6583 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6642
Qy 1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
Db 6643 ACGTGAAAGTTACACCAGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6702
Qy 1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
Db 6703 AAGCCGCAGAACCCCTGGCGACTGCTTACTTATGCGGGATTCACCGGGAATTAGTGCGTA 6762
Qy 1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
Db 6763 GGCTTACGGCCGTCTTGCTTCCAAACATTCACACGCTTTTTGACATGTCGGCGGAGGATT 6822
Qy 1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
Db 6823 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGCGACCCGGTACTGGAGACGGATATCG 6882
Qy 1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
Db 6883 CATCATTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGTCTGATGATCTTGGAGG 6942
Qy 1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
Db 6943 ACCTGGGTGTGGATCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 7002
Qy 1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
Db 7003 CCACCCATCTACCTACGGGTACTCGTTTTAAATTCGGGGCGATGATGAAATCCGGAATGT 7062
Qy 1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
Db 7063 TCCTCACACTTTTTGTCAACACAGTTTTGAATGTCGTTATCGCCAGCAGAGTACTAGAAG 7122
Qy 1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
Db 7123 AGCGGCTTAAAACGTCCAGATGTGCAGCGTTCATTGGCGACGACAACATCATACATGGAG 7182
Qy 2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
Db 7183 TAGTATCTGACAAAGAAATGGCTGAGAGGTGCGCCACCTGGCTCAACATGGAGGTTAAGA 7242
Qy 2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139
Db 7243 TCATCGACGCAGTCATCGGTGAGAGACCACCTTACTTCTGCGGCGGATTTATCTTGCAAG 7302
Qy 2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199
Db 7303 ATTCGGTTACTTCCACAGCGTGCCGCGTGGCGGACCCCCTGAAAAGGCTGTTTAAGTTGG 7362
Qy 2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
Db 7363 GTAAACCGCTCCCAGCCGACGACGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 7422
Qy 2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319
Db 7423 AAACAAAGGCGTGGTTTAGAGTAGGTATAACAGGCACTTTAGCAGTGGCCGTGACGACCC 7482
Qy 2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379
Db 7483 GGTATGAGGTAGACAATATTACACCTGTCCTACTGGCATTGAGAACTTTTGCCCAGAGCA 7542
Qy 2380 TTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGG 2439
Db 7543 AAAGAGCATTCCAAGCCATCAGAGGGGAAATAAAGCATCTCTACGGTGGTCCTAAATAGT 7602
Qy 2440 TGACGTAGTAGACACGCACCTAC----------CCACCGGCAGAATGTTTCCATACC 2486
Db 7603 CAGCATAGTACATTTCATCTGACTAATACTACAACACCACCACCATGAATAGAGGATTCT 7662
Qy 2487 CTCAGCTGAACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACC 2546
Db 7663 TTAACATGCTCGGCCGCCGCCCCTTCCCGGCCCCCACTGCCATGTGGAGGCCGCGGAGAA 7722
Qy 2547 CTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTA 2606
Db 7723 GGAGGCAGGCGGCCCCGATGCCTGCCCGCAACGGGCTGGCTTCTCAAATCCAGCAACTGA 7782
Qy 2607 GGAGGTCGATAGTCAACTTGACTTTCAAACAACGATCACCTAATCCGCCGCCAGGTCCA- 2665
Db 7783 CCACAGCCGTCAGTGCCCTAGTCATTGGACAGGCAACTAGACCTCAACCCCCACGTCCAC 7842
Qy 2666 -----CCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGA 2720
Db 7843 GCCCGCCACCGCGCCAGAAGAAGCAGGCGCCCAAGCAACCACCGAAGCCGAAGAAACCAA 7902
Qy 2721 AGAAGCAGCAAGCCAAGAGGACGAAAC---GCAAGCCTAAACCAGGGAAACGACAACGTA 2777
Db 7903 AAACGCAGGAGAAGAAGAAGAAGCAACCTGCAAAACCCAAACCCGGAAAGAGACAGCGCA 7962



```
Db      10123 AAAGGGCAGGGTATGCCCCGCTCAATTTGGAGATCACTGTCATGTCCTCGGAGGTTTTGC 10182

Qy       4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
              |||| || ||| | |||||| | ||||||||||||  ||| ||  | || || ||| ||
Db      10183 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCAAAAA 10242

Qy       5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
              | |||||||||| ||| | || ||  || |  || |  | || || || || |||   |
Db      10243 TCAAATGCTGCGGCTCCTTGGAATGTCAGCCGGCCGCTCATGCAGACTATACCTGCAAGG 10302

Qy       5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
              | || || || || ||||| || |||||||||||| || ||||| || || ||||||||||
Db      10303 TCTTCGGAGGGGTCTACCCCTTTATGTGGGGAGGAGCGCAATGTTTTTGCGACAGTGAGA 10362

Qy       5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
              |||  ||  ||||||||||||||||||  |  |  |||| ||| |    || |||||
Db      10363 ACAGCCAGATGAGTGAGGCGTACGTCGAACTGTCAGCAGATTGCGCGTCTGACCACGCGC 10422

Qy       5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
              ||  | || || ||||| || ||  ||||||| || |||||||||| ||||| ||||
Db      10423 AGGCGATTAAGGTGCACACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10482

Qy       5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
              | |||      ||| |||  ||  || || || ||||| |||||  | ||    ||| |||
Db      10483 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACTTGA 10542

Qy       5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
              | |||||||| || || || |||||  | ||| | || || || |||||||||| |||
Db      10543 AAGTCATAGCTGGACCAATTTCAGCATCGTTTACGCCATTCGATCATAAGGTCGTTATCC 10602

Qy       5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
              || || || |||||||| |||||||| || |||||||| ||||||||||||||||
Db      10603 ATCGCGGCCTGGTGTACAACTATGACTTCCCGGAATATGGAGCGATGAAACCAGGAGCGT 10662

Qy       5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
              | || || ||||||||   |||| |   |   | ||  |  | ||| |||| |||||
Db      10663 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAGGATCTCATCGCCAGCACAGACATTA 10722

Qy       5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
              |||| || |||||||| | ||||||| | || ||||| ||||| || || | |||||| |
Db      10723 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCAGGAT 10782

Qy       5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
              ||| |||||||| |||||||||||| || || ||||| ||||| ||||| || || ||||
Db      10783 TTGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCACCTTTCGGGTGTA 10842

Qy       5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
              | |||| |||  | || || ||||||     ||||| | |||||| |||| || || || |
Db      10843 AGATTGCAGTAAATCCGCTCCGAGCGGTGGACTGTTCATACGGGAACATTCCCATTTCTA 10902

Qy       5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
              ||||||||||  | || || ||| | ||  |||||||  |||||    | | |  ||| |
Db      10903 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10962

Qy       5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
              ||   |||    || |||| |||||| |||||||| || ||        ||  | ||||
Db      10963 AATGTGAAGTCAGTGAGTGCACTTATTCAGCAGACTTCGGCGGGATGGCCACCCTGCAGT 11022

Qy       5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
              |   | | ||| | || || || || || || || || || || | |||||   | |
Db      11023 ATGTATCCGACCGCGAAGGTCAATGCCCCGTACATTCGCATTCGAGCACAGCAACTCTCC 11082

Qy       5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
              | ||  ||||   ||||||    |    |||     | ||| |||| ||||||||  |||
Db      11083 AAGAGTCGACAGTACATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTTAGCACCGCGA 11142

Qy       5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
              | ||||| || || ||||| || ||||| || || |||||| | || ||||||| ||||
Db      11143 GTCCACAGGCGAACTTTATCGTATCGCTGTGTGGGAAGAAGACAACATGCAATGCAGAAT 11202

Qy       6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
              |||||||||| || ||||| ||  |  |    || || |||   |||||||||| || |
Db      11203 GTAAACCACCAGCTGACCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTTCAAG 11262

Qy       6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
              | ||  | || |||||||| |||| ||||| |||| || ||  || || || || | |
Db      11263 CCGCCATCTCAAAAACATCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11322

Qy       6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
              | |  || ||||||||||  | || |  ||||||   ||||| |  | |||||| ||||
Db      11323 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11382

Qy       6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
              |||  |          || ||| |  | | | ||||||||||||||||||||||||  |  ||
Db      11383 GACCGCTACGCCCCAATGATCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11442

Qy       6242 GCATAATGCCACGCGCCGCT 6261
              |||||||||  |  || | |
Db      11443 GCATAATGCATCAGGCTGGT 11462
```

Search completed: November 16, 2003, 05:39:13
Job time : 22826 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 12:03:02 ; Search time 8996 Seconds
(without alignments)
13508.491 Million cell updates/sec

Title: US-10-023-649-1_COPY_1_5000
Perfect score: 5000
Sequence: 1 accctacaaactaatcgatc.....ataattctagaaacgccacc 5000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	49.8	1.0	885	13	BX425603	BX425603 BX425603
2	47	0.9	1201	13	BX381961	BX381961 BX381961
c 3	43.8	0.9	1098	13	BX436460	BX436460 BX436460
c 4	42	0.8	1201	13	BX381961	BX381961 BX381961
c 5	41.2	0.8	646	14	CD393733	CD393733 Gm_ck1332
c 6	40.8	0.8	398	12	BG984025	BG984025 IL5-CN006
7	40.8	0.8	971	13	BU173419	BU173419 AGENCOURT
8	40.8	0.8	1101	29	CNS017SY	AL108460 Drosophil
9	40.6	0.8	380	12	BJ198034	BJ198034 BJ198034
10	40.6	0.8	453	12	BJ198451	BJ198451 BJ198451
11	40.6	0.8	455	12	BJ198768	BJ198768 BJ198768
12	40.6	0.8	456	12	BJ203653	BJ203653 BJ203653
13	40.4	0.8	715	12	BI195766	BI195766 602754145
c 14	40.4	0.8	1201	13	BX376097	BX376097 BX376097
15	40.2	0.8	634	10	BE017217	BE017217 fk78b02.y
16	39.8	0.8	413	14	CB384158	CB384158 TgESTzyh5
17	39.8	0.8	916	10	BG541362	BG541362 602569867
18	39.6	0.8	343	9	AW501797	AW501797 UI-HF-BR0
19	39.6	0.8	402	9	AW501918	AW501918 UI-HF-BR0
20	39.6	0.8	449	9	AW499678	AW499678 UI-HF-BR0
21	39.6	0.8	480	9	AW499647	AW499647 UI-HF-BR0
22	39.6	0.8	491	9	AW502451	AW502451 UI-HF-BR0
23	39.6	0.8	859	14	CA971504	CA971504 AGENCOURT
24	39.6	0.8	1011	10	BE563150	BE563150 601335780
25	39.4	0.8	504	13	BU926381	BU926381 sas90a08.
26	39.4	0.8	597	12	BM308456	BM308456 sak46c12.
27	39.4	0.8	619	13	BU927494	BU927494 sat13b07.
28	39.2	0.8	999	12	BM477724	BM477724 AGENCOURT
29	39	0.8	243	14	CB391641	CB391641 OSTF156A8
c 30	39	0.8	560	28	AZ738168	AZ738168 RPCI-24-1
31	39	0.8	599	13	BQ873224	BQ873224 QGI17P07.
32	39	0.8	626	13	BQ849532	BQ849532 QGB10C11.
c 33	39	0.8	997	29	CNS005TE	AL060767 Drosophil
34	39	0.8	1240	10	BG739729	BG739729 602630430
35	38.8	0.8	342	12	BI177442	BI177442 EST518387
c 36	38.8	0.8	650	28	BH900647	BH900647 Otb00189
37	38.8	0.8	687	12	BM406647	BM406647 EST580974
c 38	38.8	0.8	705	12	BI432608	BI432608 EST535369
39	38.8	0.8	718	10	BG598366	BG598366 EST503266
c 40	38.8	0.8	1101	29	CNS017GS	AL108022 Drosophil
c 41	38.8	0.8	1182	9	AL513775	AL513775 AL513775
42	38.4	0.8	413	13	BQ281915	BQ281915 WHE3026_H
43	38.4	0.8	1003	29	CNS04H1P	AL290374 Tetraodon
c 44	38.4	0.8	1076	29	CNS05HXN	AL338180 Tetraodon
c 45	38.4	0.8	1163	29	CC290181	CC290181 CH261-100

ALIGNMENTS

RESULT 1
BX425603/c
LOCUS BX425603 885 bp mRNA linear EST 15-MAY-2003
DEFINITION BX425603 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone CL0BB022ZA07 3-PRIME, mRNA sequence.
ACCESSION BX425603
VERSION BX425603.1 GI:30770486
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 885)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CL0BB022ZA07FP1.
FEATURES Location/Qualifiers
source 1. .885
/organism="Homo sapiens"

```
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CL0BB022ZA07"
                     /tissue_type="NEUROBLASTOMA"
                     /clone_lib="Homo sapiens NEUROBLASTOMA"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
BASE COUNT       15 a     23 c     69 g    154 t    624 others
ORIGIN

  Query Match             1.0%;  Score 49.8;  DB 13;  Length 885;
  Best Local Similarity   2.7%;  Pred. No. 0.012;
  Matches   15;  Conservative  169;  Mismatches  378;  Indels    0;  Gaps    0;

Qy       1058 CAACTGACGTTAGTGTGGATGACGCACAAAAACTATTGGTTGGGCTCAACCAAAGGATTG 1117
              |:    ::                                    :       ::::     : : :
Db        865 CMNNMMMMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNKNNNNNNNMMMMNNNNNKNKNK 806

Qy       1118 TCGTCAATGGTAGGACGCAAAGAAATACTAACACAATGCAGAACTATCTATTACCAGTGG 1177
               :  ::                                 :::::  :: ::: : : :  ::::
Db        805 NMMMMMNNNNNNNNNNNMNNNNNNNNNNNNNNNNMMMMMMMMMMMMMMMMMMMMMMMMM 746

Qy       1178 TCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATCGTGCCGACTTGGACGACGAGAAAG 1237
               : ::::       :::   : : :  :::: :   ::         ::     :
Db        745 MMMMMMMNNNNNNMMMMNNNKNKNKKNKKMMNMNNNKKNNNNNNNNNNNNNNNKMMNNNK 686

Qy       1238 AACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCTGCTGCTGGGCTTTCAAGACCCAGA 1297
                  ::::: |:: :      :          :       :   ::: ::::: :
Db        685 KNKGKNKKKKKKGKKGKKNTNKNMNNNNNNNNNNNNMMNNNNNNMMMMMMMMMMMMM 626

Qy       1298 AAATCACATCCATCTACAAGAAGCCTGGTACGCAAACAATTAAGAAAGTACCTGCCGTCT 1357
               : | :: | ::: : :::: :: ::                              ::
Db        625 MNAMMMNAMMMMAMMMMMMMMMMMMMMNNNNNNNNNNNNNNNNNNNNNNNNNKKKKNKNNNN 566

Qy       1358 TTGACTCATTTGTGATTCCACGCCTTACCAGCCACGGGCTCGATATGGGCTTCCGCCGTA 1417
                  ::          :  ::  :::: ::::   : : | :    :  :: ::
Db        565 NNNNNNMMNNMNNNNNNNNMAMMMMMMMMMMMMMMMMMMMMAMMMMMMMMMMMNNN 506

Qy       1418 GGCTCAAGCTGCTGCTTGAACCAACTGTCAAACCCGCACCGGCTATTACAATGGCCGATG 1477
                     :        :::   :|::::: ::::   : :  ::::   :: : :
Db        505 NNNNNNNNNNNNNKNNNNKNNNMMMMMAMAMMMMMMMMMNCMMMMMMMMMMMMMNMNN 446

Qy       1478 TGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAGAAGTGGCTGCAGCGGAAGAGATCA 1537
                 : :::::   : :       :::                |
Db        445 NNNMNMMKMKNNNKNKNNNNNNMMMMNNNNNNNNNTKNNTNNNNTNNNNTNNNNNNNNN 386

Qy       1538 GAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAAAAGAGACCGTAGAGGCAGAAGTAG 1597
                     : ::::   : :::  : : | ::||| : :::   | :  :: ::  :
Db        385 TTNNNMNNNMMMNMMMMAMMMMMMMMAMCMMAMMMAAANMMMMMNNANMMMMMMMMMMM 326

Qy       1598 ACCTCATTATGCAAGAGGCAGG 1619
               :       :   |      ||
Db        325 MNNNNNNNNCSNNNGNNNKNGG 304


RESULT 2
BX381961
LOCUS       BX381961               1201 bp    mRNA    linear   EST 08-MAY-2003
DEFINITION  BX381961 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI072YF05 3-PRIME, mRNA sequence.
ACCESSION   BX381961
VERSION     BX381961.1  GI:30453007
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI072CC03NP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI072YF05"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT       95 a    191 c    115 g     55 t    745 others
ORIGIN

  Query Match             0.9%;  Score 47;  DB 13;  Length 1201;
  Best Local Similarity   5.5%;  Pred. No. 0.1;
  Matches   38;  Conservative  251;  Mismatches  404;  Indels    2;  Gaps    1;

Qy        743 CGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGCATCTTCCAA 802
              :  : :::       |   :       ::  : :       :: : :  :        ::| :
Db        454 MNNKNKKKNNKNNNNTKNNMNNNNCNKMNNKNKKNNNNNKMNKNMMNKNNNCNNNKMCKM 513

Qy        803 ACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTGTCAGCT 862
              :    : ::: :        ::  :      ::: :   :  :          :     :
Db        514 MNMMNKMMMMNKMMNNNCNKMMNAMNNKKMNKMMMNKNNMMNCKTNNNNNNNNKNNNNNK 573

Qy        863 GTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAGTTGAGA 922
                 ::::  :          :    | :          :   |  :  :::: :
Db        574 NNNNNKKKKKKKNNNNNKNNNMNNNNANKNNNNNNNNNKNNCMNKKKKKKMKMMMMNKMN 633

Qy        923 ACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATACGCTGC 982
              :  :        : ::   . :                          :    : :
Db        634 NNNNNKKNKNNNKNNNNMNKMNNNNNKNNNMNNNNNNNNNNNNNNNKKKKKKKKNNNNNNN 693

Qy        983 GCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCGATCAGA 1042
                  :::   :: ::| : ::   ::      :          ::    |  :
Db        694 NNNNNNMKMNNNNKKNKKTNKKKKNNAKKNNTNNKMMNNKNNNNNNMMNNCNCNKKKNNK 753

Qy       1043 TGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTATTGGTTGGGC 1102
              | :          ::     :           ::   : |:   :   :| : ::::: ::
Db        754 TNMNNMMNNNNNKKNNNNNKNNNNNNNNNNNKKMKMSKCKKKKMCCKMCCMCKKKKKMBKG 813

Qy       1103 TCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAATGCAGAACT 1162
              ::: :| :: : ::  | ::::  :   ::| :    ::: :   ::::   : :
Db        814 KMVCMCKMMMKNKKCMCMKBMMMCKMCMKMCMBKKNMMMMMMKMKMMMMNKNKAMMKKDN 873

Qy       1163 ATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATCGTGCCGACT 1222
              : : :  |::    ::  ::: : |  :: ||: :::: :   : ::: :: :|: . :
Db        874 MMMMMMCAMMCMMMKKNMMMBKKNCMMKKMCCMMKKKKHKKNMMKKMMMCKKBMMCKKNK 933

Qy       1223 TGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCTGCTGCTGGG 1282
              : : : :|::   | :: : : | ::: :: :|  :: : : ::::  : :::
Db        934 MMNKMMVKKKKAMMNBKCVMKMKMKMCCKKVMMBKMCAKKKKMCMCVKKVMCMCDBKMCM 993

Qy       1283 CTTTCAAGACCCAGAAAATCACATCCATCTACAA--GAAGCCTGGTACGCAAACAATTAA 1340
              |:::::  ::||     ::: ::||  :: ::|:::   :::::     :: ::   :: :
Db        994 CKKKBMKKVCCKKMCMMMMMCAMKVMMMBAMMMMBAMVVMMMMNNMMMMCKKMVMKKK 1053

Qy       1341 GAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCACGGGCTCGA 1400
              ::  :: ::::::   :  :::::   :|  :   |:|: |   :   |:::|::  |:
Db       1054 KMCKKKBMSKKMMMNKKVCKMBBMCMSKGGCMMCGCMCKGCVGCMNSCCMMBGSKGCCKK 1113

Qy       1401 TATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTG 1435
```

```
Db        1114 KMBKCMMCCCMCMCBCMCBKKVCVKKKKKCKKKKKK 1148


RESULT 3
BX436460/c
LOCUS       BX436460               1098 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX436460 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP001YL10
            5-PRIME, mRNA sequence.
ACCESSION   BX436460
VERSION     BX436460.1  GI:30779481
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1098)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0CAP001DF05QP1.
FEATURES             Location/Qualifiers
     source          1. .1098
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0CAP001YL10"
                     /tissue_type="THYMUS"
                     /clone_lib="Homo sapiens THYMUS"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
BASE COUNT       56 a    289 c    225 g    229 t    299 others
ORIGIN

  Query Match             0.9%;  Score 43.8;  DB 13;  Length 1098;
  Best Local Similarity   20.5%;  Pred. No. 0.87;
  Matches   83;  Conservative  130;  Mismatches  192;  Indels    0;  Gaps    0;

Qy       2667 GCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGAGGATGGGTGAAACAGCTACAGAT 2726
Db        682 GGCGGACNMMGGGMMMMMAARAAAANANGKAGDAAAKKAAKAKKDAMAAMAAKAAKAAA 623

Qy       2727 TGACTACAAAAATCACGAAATCATGACTGCGGCTGCATCGCAAGGACTTACGCGGAAAGG 2786
Db        622 AAKDKAKDAAARKKGKKRAAKKKAGKAAKAAKAAKAAADAAAADAAKAKAAGKKKAAAAA 563

Qy       2787 CGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCACTCTACTCGCAGACTTCTGAGCA 2846
Db        562 RKDDKKKKNAAAAAAAAAAAAAKARKKKKAKKKAAARKKAGKKAMAAAAVKKKKKAAAAM 503

Qy       2847 CGTGAACGTGTTACTTACACGCACAGAAAAACGCATTGTCTGGAAGACGCTAGCTGGTGA 2906
Db        502 GAAAAAAAMKKKAAAAAMAKKAAMMKAAMMMAMMKAKKAAKMMMAMAKMAAKAMKKMKK 443

Qy       2907 TCCCTGGATAAAGACACTTACAGCTAAATATCCCGGGGATTTCACGGCTTCATTGGACGA 2966
Db        442 MMMMMMKAMAMMKAMMMAKAMKMMMMMMKMMMMKMMKMKKMVMCACMCMMVKMMKMMAA 383

Qy       2967 CTGGCAGCGCGAACACGACGCCATTATGGCACGCGTTCTTGATAAGCCGCAGACAGCTGA 3026
Db        382 MMAVKMVMKKCCAKKAKACCCMCAACACKAKCKCCCACAAKAKKAKCCKCAAMAAAAACA 323

Qy       3027 TGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAGGCTTTAGA 3071
```

```
Db         322 AKCAAKVKAKASAAAAAKKAKKKRKGGAGGAAGAAAKGAKTAKKA 278


RESULT 4
BX381961/c
LOCUS       BX381961               1201 bp    mRNA    linear   EST 08-MAY-2003
DEFINITION  BX381961 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI072YF05 3-PRIME, mRNA sequence.
ACCESSION   BX381961
VERSION     BX381961.1  GI:30453007
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI072CC03NP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI072YF05"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT       95 a    191 c    115 g     55 t    745 others
ORIGIN

  Query Match             0.8%;  Score 42;  DB 13;  Length 1201;
  Best Local Similarity   4.7%;  Pred. No. 3.2;
  Matches   32;  Conservative  241;  Mismatches  398;  Indels    4;  Gaps    2;

Qy        810 CCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTGTCAG---CTGTGA 866
Db       1071 BMMNKKKMMSKVMMMGKMMMMKBKMMGKKKKKNNKKKKBBKTVKKKKTVKKKBMKTGKKK 1012

Qy        867 AGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAGTTGAGAACTT 926
Db       1011 KKKGKMMGGBMMKVMMMGKGKMVHGKGKBMMBGKGKMMMMTGKMVKKBMMGGKMKMKMKB 952

Qy        927 GGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATACGCTGCGCGG 986
Db        951 GMVNKKTMMMMBKKMNKKMNMMGKKVMMGKKKMMKKNMMDMMMMKKGGKMMKKGNMMVKK 892

Qy        987 CGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCGATCAGATGAC 1046
Db        891 KNMMKKKGKKTGKKKKKKNHMMKKTMNMNKKKKMKMKKKKKKNMMVKGKMKGKMGKKKVM 832

Qy       1047 AGGGATTCTGGCAACTGACGTTAGTGTGGATGACG-CACAAAAACTATTGGTTGGGCTCA 1105
Db        831 KGKGMMNMKKKMGKGBKMCMVKMMMMMGKGGKMGGKMMMMGMSKMKMNNNNNNNNNNMN 772

Qy       1106 ACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAATGCAGAACTATC 1165
Db         771 NNNNMMNNNNKKNNKNAMNNMMMNGNGNNKKNNNNNNMNNKKMNNANNMMTNNMMMMNA 712

Qy       1166 TATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATCGTGCCGACTTGG 1225
```

```
Db        711 MMNMMNNNNKMKNNNNNNNNNNNNNMMMMMMMNNNNNNNNNNNNNNKNNNMNNNNNKMN 652

Qy       1226 ACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCTGCTGCTGGGCTT 1285

Db        651 KNNNNMNNNMMMNNNNNNKMNKKKKMKMMMMMNKGNNMNNNNNNNNNNMNTNNNNKNNN 592

Qy       1286 TCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCTGGTACGCAAACAATTAAGAAAG 1345

Db        591 MNNNNNMMMMMMMNNNNNNMNNNNNNMNNNNNNNNAMGNKKNNMNKKKMNKMMNNKTNKKMN 532

Qy       1346 TACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCACGGGCTCGATATGG 1405

Db        531 GNNNKKMNKKKKMNKKNKKMGKMNNNGNNNMNKKNMNKMNNNNNMMNMNNKMNGNNNNKN 472

Qy       1406 GCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACTGTCAAACCCGCACCGGCTATTA 1465

Db        471 NMANNNNMNNMMMNMNNKKNNTNGNNNNNMMMMMMMNKNMMNKKMMMNMKGMMMMNN 412

Qy       1466 CAATGGCCGATGTGG 1480

Db        411 NKKKMGMNNNNNTNG 397
```

RESULT 5
CD393733/c
LOCUS CD393733 646 bp mRNA linear EST 01-JUN-2003
DEFINITION Gm_ck13327 Soybean induced by Salicylic Acid Glycine max cDNA 3', mRNA sequence.
ACCESSION CD393733
VERSION CD393733.1 GI:31308530
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 646)
AUTHORS Tian,A.-G., Wang,J., Cui,P., Han,Y.-J., Xu,H., Cong,L.-J., Huang ,X.-G., Wang,X.-L., Jiao,Y.-Z., Wang,B.-J., Wang,Y.-J., Zhang,J.-S. , Chen,S.-Y. and Yu,J.
TITLE Soybean Expressed Sequence Tags Sequencing
JOURNAL Unpublished
COMMENT Contact: Chen S-Y
Plant Biotechnology Laboratory
Institute of Genetics and Developmental Biology, CAS, China
Datun road, Beijing 100101, China
Tel: 86-10-64886859
Fax: 86-10-64873428
Email: sychen@genetics.ac.cn
Email: sychen@genetics.ac.cn
Seq primer: T7 primer.
FEATURES Location/Qualifiers
source 1. .646
/organism="Glycine max"
/mol_type="mRNA"
/cultivar="Kefeng 1"
/db_xref="taxon:3847"
/tissue_type="Seedlings"
/dev_stage="two-week seedlings"
/lab_host="XL1-Blue MRF' strain"
/clone_lib="Soybean induced by Salicylic Acid"
/note="Vector: pBluescript SK+; Site_1: EcoR I; Site_2: Xho I; The cDNA library was constructed by He, C-Y from mRNA isolated from two-week seedlings (cultivar Kefeng 1) treated by spraying 2.0mM salicylic acid for 24, 36, 48 and 72 h. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI striction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL1-Blue MRF' host cells (Stratagene)."
BASE COUNT 206 a 134 c 134 g 172 t
ORIGIN

```
  Query Match              0.8%;  Score 41.2;  DB 14;  Length 646;
  Best Local Similarity   48.9%;  Pred. No. 3.4;
  Matches  139;  Conservative    0;  Mismatches  143;  Indels    2;  Gaps    1;

Qy       2331 AGTGCTTCTAAATGGGGTTAAGCACCCCGTTAACACTCTGTACATTGATGAGGCATTTGC 2390

Db        599 AGTTCTGCCAAATCAAGAGGGGTTTTAAGATCATAGTGTGTGTTTTGCAACAATGGCTAA 540

Qy       2391 CTGCCATGCAGGGACGCTGCTGGCACTGATTGCCATCGTCAAACCTAAGAAAGTGGTATT 2450

Db        539 GTTCTTTGCTGCTATGATCTTGGCACTCTTTGCCATTTCCATACTTCAAACAGTGGTAAT 480

Qy       2451 GTGCGGGGACCCAAAACAATGCGGCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAA 2510

Db        479 GGCTGCTAATGAACAAGGAGGCCACTTGTATGACAACAAGAGCAAATATGGAAGTGGAAG 420

Qy       2511 CCATGACATATGCACTGAAGTGTACCATAAAAGCATCTCTAGGAGGTGCACACAGACTGT 2570

Db        419 TGTCAAGAGTTACCGTAAAATGCCCATCACAATG--CTCGAGGAGATGTAGCCAGACCCA 362

Qy       2571 AACCGCCATCGTCTCCACGCTCTTCTACGACAAGCGAATGAAGA 2614

Db        361 ATACCACAAGCCCTGCATGTTTTTCTGTCAGAAGTGCTGCAGGA 318
```

RESULT 6
BG984025/c
LOCUS BG984025 398 bp mRNA linear EST 12-JUN-2001
DEFINITION IL5-CN0065-190301-403-d07 CN0065 Homo sapiens cDNA, mRNA sequence.
ACCESSION BG984025
VERSION BG984025.1 GI:14386760
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 398)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=IL5&t2=IL5-CN0065-190301-403-d07&t3=2001-03-19&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 370.
FEATURES Location/Qualifiers
source 1. .398
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="CN0065"

/note="Organ: colon_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 34 a 156 c 97 g 111 t
ORIGIN

Query Match 0.8%; Score 40.8; DB 12; Length 398;
Best Local Similarity 51.7%; Pred. No. 3.1;
Matches 93; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 1445 TCAAACCCGCACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAG 1504
Db 243 TTAAAACAGCAGCGGCAACTGCAGCAGCAGCAGCAGCGAGACCCCGAGGCACACATCAAA 184

Qy 1505 AAGCTGAAGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTG 1564
Db 183 CACCTGCTGCACCAGCGGCAGCGGCGCATAGAGGAGCAGAAGGAGGAGCGGCGCCGCGTG 124

Qy 1565 AAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGAGCAG 1624
Db 123 GAGGAGCAACAGCGGCGGGAGCGGGAGCAGCGGAAGCTGCAGGAGGAGGAGCAGCAGCGG 64

RESULT 7
BU173419
LOCUS BU173419 971 bp mRNA linear EST 04-SEP-2002
DEFINITION AGENCOURT_7942821 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6171839 5', mRNA sequence.
ACCESSION BU173419
VERSION BU173419.1 GI:22687403
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 971)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13540 row: m column: 24
High quality sequence start: 106
High quality sequence stop: 362.
FEATURES Location/Qualifiers
source 1. .971
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6171839"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_67"
/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Library constructed by Life Technologies."
BASE COUNT 269 a 276 c 308 g 118 t
ORIGIN

Query Match 0.8%; Score 40.8; DB 13; Length 971;
Best Local Similarity 51.7%; Pred. No. 6.2;
Matches 93; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 1445 TCAAACCCGCACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAG 1504
Db 329 TTAAAACAGCAGCAGCAGCTGCAACAGCAGCAGCAGCGAGACCCCGAGGCACACATCAAA 388

Qy 1505 AAGCTGAAGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTG 1564
Db 389 CACCTGCTGCACCAGCGGCAGCGGCGCATAGAGGAGCAGAAGGAGGAGCGGCGCCGCGTG 448

Qy 1565 AAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGAGCAG 1624
Db 449 GAGGAGCAACAGCGGCGGGAGCGGGAGCAGCGGAAGCTGCAGGAGAAGGAGCAGCAGCGG 508

RESULT 8
CNS017SY
LOCUS CNS017SY 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC BACN37L08 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL108460
VERSION AL108460.1 GI:5628764
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN37L08"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : SP6"
BASE COUNT 254 a 176 c 160 g 152 t 359 others
ORIGIN

Query Match 0.8%; Score 40.8; DB 29; Length 1101;
Best Local Similarity 16.4%; Pred. No. 6.9;
Matches 65; Conservative 169; Mismatches 160; Indels 2; Gaps 1;

Qy 1269 GGGCTGCTGCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCTGGTAC 1328
Db 665 RSGSGGSSCCGGKKAKGGVRGGRVCCAGGGASCACMAADCGGCCAKMACCSSSSASSSGS 724

Qy 1329 GCAAACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAG 1388
Db 725 SCASTSSSASRGGMVSSCACSGSGSGGASACGASGAGGGSGRGGGSGCCASSGVCGAACS 784

Qy 1389 CCACGGGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACTGTCAA 1448
Db 785 SASCSSASMCGVSSGSSCSASCGSCCGVSSCSAVSASSASSVMSKVASAVASCSAVASGM 844

Qy 1449 ACCCGCACCGGCTATTACAATGGCCGATGTGG--AGCATCTGCGTGGCTTACAGCAAGAA 1506

```
Db        845 SAGAVSSSCRSSVASSVSAAASVSSSVSSSSSSSVVSAAAVSASSSSASSASMAVAAAAA 904

Qy       1507 GCTGAAGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAA 1566
              :  : :::|: ::: :: :::: |::::   :|  |:::: ::: :::| :: ::: : |
Db        905 VAVSSVSVASVVSSSSSCSSSSSASVVVSASVAASASSVSSSSSSSVSTSSSASVSSVSA 964

Qy       1567 ATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGAGCAGGT 1626
              : ::|::  :|: :::  : :::|:||: | |    | | :: |:|: :::  :: ::
Db        965 VSMSAVVSSSASSSSSVSVSSVVAVAASAAAAAAAAAAAAASSSSASAVAVVSASSSSSSA 1024

Qy       1627 AGCGTGGAGACACCACGAGGACACATCAGGGTGACA 1662
                ::: ::|: :::::::: :  :::: : ::: :|:|
Db       1025 SSSSSSSASSSVSSVSSSSCSVVSVVSSSSVSVAVA 1060
```

RESULT 9
BJ198034
LOCUS BJ198034 380 bp mRNA linear EST 24-JAN-2002
DEFINITION BJ198034 normalized full length cDNA library, chloronemata, caulonemata and rhizoid-like protonemata Physcomitrella patens subsp. patens cDNA clone pphn30g20 5', mRNA sequence.
ACCESSION BJ198034
VERSION BJ198034.1 GI:18365956
KEYWORDS EST.
SOURCE Physcomitrella patens subsp. patens
ORGANISM Physcomitrella patens subsp. patens
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta; Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE 1 (bases 1 to 380)
AUTHORS Fujita,T., Shin-i,T., Seki,M., Kamiya,A., Uchiyama,I., Nishiyama,T., Carninci,P., Hayashizaki,Y., Shinozaki,K., Kohara,Y. and Hasebe,M.
TITLE Comparison of the moss Physcomitrella patens genome with flowering plants genome
JOURNAL Unpublished
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
A backbone of the vector is pBluescript II, that was in vivo excised from a modified lPS phage vector (Mo bi Tec, Germany). XhoI digested-5' end of cDNA is ligated to SalI site of the vector, and the BamHI digested-3' end including poly-A tail is ligated to BamHI site of the vector. cDNA instert could be amplified with conventional T7 and T3 primers. This normarized full- length cDNA library was generated basically according to the method described in Genome Research 10, 1617-1630 (2000), Carninci, P. et al. Protonemata were blended by the POLYTRON, and then cultivated on the BCD medium containing 1uM NAA (naphthalene acetic acid) for 8 to 11 days under the continuous light.
FEATURES Location/Qualifiers
source 1. .380
/organism="Physcomitrella patens subsp. patens"
/mol_type="mRNA"
/sub_species="patens"
/db_xref="taxon:145481"
/clone="pphn30g20"
/tissue_type="mixture of chloronemata, caulonemata and rhizoid-like protonemata"
/clone_lib="normalized full length cDNA library, chloronemata, caulonemata and rhizoid-like protonemata"
BASE COUNT 131 a 75 c 121 g 53 t
ORIGIN

Query Match 0.8%; Score 40.6; DB 12; Length 380;
Best Local Similarity 50.8%; Pred. No. 3.4;
Matches 97; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

```
Qy       1474 GATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAGAAGTGGCTGCAGCGGAAGAG 1533
              || | |||| |       |  ||||  | || || ||||||||  |  | ||  |||||
```

```
Db        103 GAGGAGGAGGAGGACGAAGATTTACGACCAGCAGTTGAAGAAGAAGAAGAAGAAGAAGAA 162

Qy       1534 ATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAAAAGAGACCGTAGAGGCAGAA 1593
                 | || |    ||    |   |   |||  |||| ||||  | ||||  | || |
Db        163 GAAGCTGTAGTCGAAGCAGAAGTCGACGAAGAAGAAGAAGAAGAAGCTGTAGTCGAAGCA 222

Qy       1594 GTAGACCTCATTATGCAAGAGGCAGGAGCAGGTAGCGTGGAGACACCACGAGGACACATC 1653
              | || |  |       |||| | || ||||||  | |  |       |.|| ||   |
Db        223 GAAGTCGACGAAGAAGAAGAAGAAGCAGCAGGAGGAGGAGGACGCATAAGAAGAAGAAGA 282

Qy       1654 AGGGTGACAAG 1664
              ||  || |||
Db        283 AGAAGGAGAAG 293
```

RESULT 10
BJ198451
LOCUS BJ198451 453 bp mRNA linear EST 24-JAN-2002
DEFINITION BJ198451 normalized full length cDNA library, chloronemata, caulonemata and rhizoid-like protonemata Physcomitrella patens subsp. patens cDNA clone pphn31f22 5', mRNA sequence.
ACCESSION BJ198451
VERSION BJ198451.1 GI:18366372
KEYWORDS EST.
SOURCE Physcomitrella patens subsp. patens
ORGANISM Physcomitrella patens subsp. patens
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta; Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE 1 (bases 1 to 453)
AUTHORS Fujita,T., Shin-i,T., Seki,M., Kamiya,A., Uchiyama,I., Nishiyama,T., Carninci,P., Hayashizaki,Y., Shinozaki,K., Kohara,Y. and Hasebe,M.
TITLE Comparison of the moss Physcomitrella patens genome with flowering plants genome
JOURNAL Unpublished
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
A backbone of the vector is pBluescript II, that was in vivo excised from a modified lPS phage vector (Mo bi Tec, Germany). XhoI digested-5' end of cDNA is ligated to SalI site of the vector, and the BamHI digested-3' end including poly-A tail is ligated to BamHI site of the vector. cDNA instert could be amplified with conventional T7 and T3 primers. This normarized full- length cDNA library was generated basically according to the method described in Genome Research 10, 1617-1630 (2000), Carninci, P. et al. Protonemata were blended by the POLYTRON, and then cultivated on the BCD medium containing 1uM NAA (naphthalene acetic acid) for 8 to 11 days under the continuous light.
FEATURES Location/Qualifiers
source 1. .453
/organism="Physcomitrella patens subsp. patens"
/mol_type="mRNA"
/sub_species="patens"
/db_xref="taxon:145481"
/clone="pphn31f22"
/tissue_type="mixture of chloronemata, caulonemata and rhizoid-like protonemata"
/clone_lib="normalized full length cDNA library, chloronemata, caulonemata and rhizoid-like protonemata"
BASE COUNT 138 a 110 c 134 g 71 t
ORIGIN

Query Match 0.8%; Score 40.6; DB 12; Length 453;
Best Local Similarity 50.8%; Pred. No. 3.9;
Matches 97; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

```
Qy       1474 GATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAGAAGTGGCTGCAGCGGAAGAG 1533
              || | |||| |       |  ||||  | || || ||||||||  |  | ||  |||||
```

```
Db        103 GAGGAGGAGGAGGACGAAGATTTACGACCAGCAGTTGAAGAAGAAGAAGAAGAAGAAGAA 162

Qy       1534 ATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAAAAGAGACCGTAGAGGCAGAA 1593

Db        163 GAAGCTGTAGTCGAAGCAGAAGTCGACGAAGAAGAAGAAGAAGAAGCTGTAGTCGAAGCA 222

Qy       1594 GTAGACCTCATTATGCAAGAGGCAGGAGCAGGTAGCGTGGAGACACCACGAGGACACATC 1653

Db        223 GAAGTCGACGAAGAAGAAGAAGAAGCAGCAGGAGGAGGAGGACGCATAAGAAGAAGAAGA 282

Qy       1654 AGGGTGACAAG 1664

Db        283 AGAAGGAGAAG 293


RESULT 11
BJ198768
LOCUS       BJ198768                455 bp    mRNA    linear   EST 24-JAN-2002
DEFINITION  BJ198768 normalized full length cDNA library, chloronemata,
            caulonemata and rhizoid-like protonemata Physcomitrella patens
            subsp. patens cDNA clone pphn32c20 5', mRNA sequence.
ACCESSION   BJ198768
VERSION     BJ198768.1  GI:18366689
KEYWORDS    EST.
SOURCE      Physcomitrella patens subsp. patens
  ORGANISM  Physcomitrella patens subsp. patens
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   Fujita,T., Shin-i,T., Seki,M., Kamiya,A., Uchiyama,I., Nishiyama,T.
            , Carninci,P., Hayashizaki,Y., Shinozaki,K., Kohara,Y. and Hasebe
            ,M.
  TITLE     Comparison of the moss Physcomitrella patens genome with flowering
            plants genome
  JOURNAL   Unpublished
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp
            A backbone of the vector is pBluescript II, that was in vivo
            excised from a modified lPS phage vector (Mo bi Tec, Germany). XhoI
            digested-5' end of cDNA is ligated to SalI site of the vector, and
            the BamHI digested-3' end including poly-A tail is ligated to BamHI
            site of the vector. cDNA instert could be amplified with
            conventional T7 and T3 primers. This normarized full- length cDNA
            library was generated basically according to the method described
            in Genome Research 10, 1617-1630 (2000), Carninci, P. et al.
            Protonemata were blended by the POLYTRON, and then cultivated on
            the BCD medium containing 1uM NAA (naphthalene acetic acid) for 8
            to 11 days under the continuous light.
FEATURES             Location/Qualifiers
     source          1. .455
                     /organism="Physcomitrella patens subsp. patens"
                     /mol_type="mRNA"
                     /sub_species="patens"
                     /db_xref="taxon:145481"
                     /clone="pphn32c20"
                     /tissue_type="mixture of chloronemata, caulonemata and
                     rhizoid-like protonemata"
                     /clone_lib="normalized full length cDNA library,
                     chloronemata, caulonemata and rhizoid-like protonemata"
BASE COUNT      138 a    111 c    134 g     72 t
ORIGIN

  Query Match             0.8%;  Score 40.6;  DB 12;  Length 455;
  Best Local Similarity   50.8%;  Pred. No. 3.9;
  Matches   97;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy       1474 GATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAGAAGTGGCTGCAGCGGAAGAG 1533

Db        103 GAGGAGGAGGAGGACGAAGATTTACGACCAGCAGTTGAAGAAGAAGAAGAAGAAGAAGAA 162

Qy       1534 ATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAAAAGAGACCGTAGAGGCAGAA 1593

Db        163 GAAGCTGTAGTCGAAGCAGAAGTCGACGAAGAAGAAGAAGAAGAAGCTGTAGTCGAAGCA 222

Qy       1594 GTAGACCTCATTATGCAAGAGGCAGGAGCAGGTAGCGTGGAGACACCACGAGGACACATC 1653

Db        223 GAAGTCGACGAAGAAGAAGAAGAAGCAGCAGGAGGAGGAGGACGCATAAGAAGAAGAAGA 282

Qy       1654 AGGGTGACAAG 1664

Db        283 AGAAGGAGAAG 293


RESULT 12
BJ203653
LOCUS       BJ203653                456 bp    mRNA    linear   EST 25-JAN-2002
DEFINITION  BJ203653 normalized full length cDNA library, chloronemata,
            caulonemata and rhizoid-like protonemata Physcomitrella patens
            subsp. patens cDNA clone pphn49j14 5', mRNA sequence.
ACCESSION   BJ203653
VERSION     BJ203653.1  GI:18372051
KEYWORDS    EST.
SOURCE      Physcomitrella patens subsp. patens
  ORGANISM  Physcomitrella patens subsp. patens
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE   1  (bases 1 to 456)
  AUTHORS   Fujita,T., Shin-i,T., Seki,M., Kamiya,A., Uchiyama,I., Nishiyama,T.
            , Carninci,P., Hayashizaki,Y., Shinozaki,K., Kohara,Y. and Hasebe
            ,M.
  TITLE     Comparison of the moss Physcomitrella patens genome with flowering
            plants genome
  JOURNAL   Unpublished
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp
            A backbone of the vector is pBluescript II, that was in vivo
            excised from a modified lPS phage vector (Mo bi Tec, Germany). XhoI
            digested-5' end of cDNA is ligated to SalI site of the vector, and
            the BamHI digested-3' end including poly-A tail is ligated to BamHI
            site of the vector. cDNA instert could be amplified with
            conventional T7 and T3 primers. This normarized full- length cDNA
            library was generated basically according to the method described
            in Genome Research 10, 1617-1630 (2000), Carninci, P. et al.
            Protonemata were blended by the POLYTRON, and then cultivated on
            the BCD medium containing 1uM NAA (naphthalene acetic acid) for 8
            to 11 days under the continuous light.
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="Physcomitrella patens subsp. patens"
                     /mol_type="mRNA"
                     /sub_species="patens"
                     /db_xref="taxon:145481"
                     /clone="pphn49j14"
                     /tissue_type="mixture of chloronemata, caulonemata and
                     rhizoid-like protonemata"
                     /clone_lib="normalized full length cDNA library,
                     chloronemata, caulonemata and rhizoid-like protonemata"
BASE COUNT      138 a    111 c    134 g     73 t
ORIGIN

  Query Match             0.8%;  Score 40.6;  DB 12;  Length 456;
  Best Local Similarity   50.8%;  Pred. No. 4;
  Matches   97;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy       1474 GATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAGAAGTGGCTGCAGCGGAAGAG 1533
```

```
Db        103 GAGGAGGAGGAGGACGAAGATTTACGACCAGCAGTTGAAGAAGAAGAAGAAGAAGAAGAA 162

Qy       1534 ATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAAAAGAGACCGTAGAGGCAGAA 1593

Db        163 GAAGCTGTAGTCGAAGCAGAAGTCGACGAAGAAGAAGAAGAAGAAGCTGTAGTCGAAGCA 222

Qy       1594 GTAGACCTCATTATGCAAGAGGCAGGAGCAGGTAGCGTGGAGACACCACGAGGACACATC 1653

Db        223 GAAGTCGACGAAGAAGAAGAAGAAGCAGCAGGAGGAGGAGGACGCATAAGAAGAAGAAGA 282

Qy       1654 AGGGTGACAAG 1664

Db        283 AGAAGGAGAAG 293


RESULT 13
BI195766
LOCUS       BI195766                715 bp    mRNA    linear   EST 10-JUL-2001
DEFINITION  602754145F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4889938 5',
            mRNA sequence.
ACCESSION   BI195766
VERSION     BI195766.1  GI:14650786
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 715)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM1766  row: i  column: 11
            High quality sequence stop: 556.
FEATURES             Location/Qualifiers
     source          1. .715
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4889938"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_19"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      202 a    194 c    251 g     68 t
ORIGIN

  Query Match             0.8%;  Score 40.4;  DB 12;  Length 715;
  Best Local Similarity   50.5%;  Pred. No. 6.5;
  Matches   98;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy       1445 TCAAACCCGCACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAG 1504

Db        223 TTAAAACAGCAGCAGCAGCTGCAGCAGCAGCAGCAGCGAGACCCCGAGGCACACATCAAA 282

Qy       1505 AAGCTGAAGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTG 1564

Db        283 CACCTGCTGCACCAGCGGCAGCGGCGCATAGAGGAGCAGAAGGAGGAGCGGCGCCGCGTG 342
```

```
Qy       1565 AAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGAGCAG 1624

Db        343 GAGGAGCAACAGCGGCGGGAGCGGGAGCAGCGGAAGCTGCAGGAGAAGGAGCAGCAGCGG 402

Qy       1625 GTAGCGTGGAGACA 1638

Db        403 CGGCCTGGAGGACA 416


RESULT 14
BX376097/c
LOCUS       BX376097                1201 bp    mRNA    linear   EST 08-MAY-2003
DEFINITION  BX376097 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
            cDNA clone CS0DC022YM12 5-PRIME, mRNA sequence.
ACCESSION   BX376097
VERSION     BX376097.1  GI:30434756
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 2866.f
            Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DC022BG06QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DC022YM12"
                     /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      79 a    152 c     66 g    270 t    634 others
ORIGIN

  Query Match             0.8%;  Score 40.4;  DB 13;  Length 1201;
  Best Local Similarity   13.7%;  Pred. No. 9.7;
  Matches  135;  Conservative  297;  Mismatches  551;  Indels    2;  Gaps    2;

Qy       1487 TGCGTGGCTTACAGCAAGAAGCTGAAGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCC 1546

Db       1186 KRKGGKGDAAADAAAAAAAAAGKAAAAAARDAKGKWAGKKAAAAGARAAKRKARAGKAGK 1127

Qy       1547 TGCCACCCTTGCTCCCTGAAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTA 1606

Db       1126 ARGAAAAAGKAGAGKAGAGGGARRKAKAGAGTARWGKTAAAGKWRKARAAKKKRKADAAD 1067

Qy       1607 TGCAAGAGGCAGGAGCAGGTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTT 1666

Db       1066 GGKAAKAGKWAGAAAGAGKAGGGKKGGGKAAAKKTMTMKTMMMMMNMMMMMMMMMMMM 1007

Qy       1667 ACCCAGGCGAAGAGAAGATTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATA 1726

Db       1006 MMMMMMMMMMHKMHMHMKMKKHMMHKTMTNTKWMKTKTMMMMMMMMMVKKMMMKMMMM 947

Qy       1727 GTGAAAAACTGGCGTGTATCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAG 1786

Db        946 KMMMMMNYMMMTGMMMMMGKHMMVMVMMMMMMMGMMMKTMMMKMMMGGKGGTMGGMGG 887
```

```
Qy       1787 GTAGGGCAGGGAGATACAAAGTCGAGCCATACCACGGTAAGGTCATTGTACCAGAAGGGA 1846
Db        886 GMMMVGMRGGGVMMMGTGKKTMGKWNMMGKMMGMMMGNTDMMKTMTMKMMMKHMTMTKGN 827

Qy       1847 CGGCGGTCCCTGTTCAAGACTTCCAGGCATTGAGTGAGAGCGCTACGATCGTTTTCAACG 1906
Db        826 GKTMHGKMMNMMVKMKTMMNVGAAAAAAADAGGGAAAAAAAAGAAAAAKAGAKAGKHMNK 767

Qy       1907 AGAGGGAGTTCGTAAACAGATACCTGCACCACATCGCAATCAACGGAGGAGCGCTAAACA 1966
Db         766 NMGKTMMGTMMMGTGKMMGKMGTMHMTMKMVMMMMMMMMMKGKMMMMKMGKMNMMAMMK 707

Qy       1967 CTGACGAAGAGTACTATAAGACTGTAAAGAC-TCAGGACACAGACTCAGAATACGTCTTC 2025
Db         706 MKMMTMTMTGTMMMTMTMMGKKKTMMTMKNBNKMMKKAGKKMNANMKKNNKNAMNNNGKM 647

Qy       2026 GATATTGACGCACGAAAGTGTGTTAAGCGAGAAGACGCAGGTCCCTTGTGCCTAACCGGT 2085
Db         646 NNMKKKTKKKNNMKKKNMKKNNNNNMMNNKMNMMKMGKMNMGKAKANNMKMKMNNNKMMMN 587

Qy       2086 GATCTGGTAGATCCACCATTTCACGAGTTTGCGTACGAGAGTCTCAAGACACGACCAGCA 2145
Db         586 NMKMKKKMKNNMKNNMKKNKKMNMNNNKKKKMMMKMNNMNGK-KMGKNNNMNAGKKMKMM 528

Qy       2146 GCACCTCACAAAGTCCCAACCATCGGAGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGA 2205
Db         527 MMMMNMNMNMMMNNMMNMNMKKNNNNMNMKKMNNMMNMMMAMAMMMGNNNGKKMNMMMKM 468

Qy       2206 ATCATCAAAAGCGCTGTGACTAAGAAAGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGC 2265
Db         467 GNGMMMNANMMMAKKNAAAMNAAANAAGKKGKNNAKAAAANAANANAKKAKAANAACNNC 408

Qy       2266 GCAGAAATCATCAGGGATGTAAGGAGGATGAGACGTATGGATGTTGCTGCTAGGACTGTC 2325
Db         407 MNMACMNNKAKNNKKKAKKNNNNNMMNMMKANKNAKAAKCNCNANNNKNAAAKKAAKNAN 348

Qy       2326 GATTCAGTGCTTCTAAATGGGGTTAAGCACCCCGTTAACACTCTGTACATTGATGAGGCA 2385
Db         347 NAANAAANNAAACNANANNNCKNNCACKAAAANNANCANNAAKAAAKNNAAAAKACAKCA 288

Qy       2386 TTTGCCTGCCATGCAGGGACGCTGCTGGCACTGATTGCCATCGTCAAACCTAAGAAAGTG 2445
Db         287 GKAKNNKKAAAAAKANANNNNCANCNANANAKNNKACANAKAACTAANANCCAKAATAAC 228

Qy       2446 GTATTGTGCGGGGACCCAAAACAAT 2470
Db         227 GTATAAAGAACCNCCACAAAACAKT 203
```

RESULT 15
BE017217
LOCUS BE017217 634 bp mRNA linear EST 30-AUG-2000
DEFINITION fk78b02.y1 Zebrafish Research Genetics C32 fin Danio rerio cDNA 5' similar to SW:APE_BRARE O42364 APOLIPOPROTEIN E PRECURSOR ;, mRNA sequence.
ACCESSION BE017217
VERSION BE017217.1 GI:8281665
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes ; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 634)
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE WashU Zebrafish EST Project 1998
JOURNAL Unpublished
COMMENT Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
cDNA Library Preparation: Ning Wu. cDNA Library Arrayed by: Research Genetics. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: Research Genetics web address: http://www.researchgenetics.com/
Seq primer: T3 ET from Amersham
High quality sequence stop: 1.
FEATURES Location/Qualifiers
source 1. .634
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/tissue_type="Fin"
/lab_host="GeneHogs (HS996, a phage-resistant isolate of DH10B)"
/clone_lib="Zebrafish Research Genetics C32 fin"
/note="Vector: pT7T3D-Pac with a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was prepared from zebrafish(C32) fin, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is non-normalized. Library was constructed by Ning Wu. NOTE: This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info.llnl.gov)for further information"
BASE COUNT 190 a 170 c 142 g 132 t
ORIGIN

Query Match 0.8%; Score 40.2; DB 10; Length 634;
Best Local Similarity 50.3%; Pred. No. 6.7;
Matches 99; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

```
Qy       4115 TGTACCGAAAATGGCCGGCTGCTTTTGATAGACAGCCAATAGCTGTCGGGACGGCTAGAC 4174
Db         60 TGGACCGATTATGGCCTGATGTGTCTTAACTCTACACACAAACTGTCGGCTTGGTGCAAA 119

Qy       4175 TTGTGAAGCACGAACCGCTCATCATACATGCTGTAGGACCCAATTTTTCTAAGATGCCGG 4234
Db        120 ACATCAAGGGCTCCCAACTCATCACAGAGCTTGAAACACTAATTACTGACACCATGGCTG 179

Qy       4235 AACCGGAGGGCGACCTTAAGCTCGCAGCTGCCTACATGAGCATAGCGTCCATCGTCAACG 4294
Db        180 AACTGAGCTCATACAGTGAAAATATACAAACCCACATGACCTCATACTCCTTTGATGCTG 239

Qy       4295 CTGAGCGGATTACAAAA 4311
Db        240 CTGGTCAGATCAGAAAA 256
```

Search completed: November 15, 2003, 19:56:26
Job time : 9001 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 22:31:43 ; Search time 12073 Seconds
(without alignments)
13055.125 Million cell updates/sec

Title: US-10-023-649-1_COPY_5000_11484
Perfect score: 6485
Sequence: 1 caccgccagcaacgacaacc.....tattttgtttttaaaatttc 6485

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	58	0.9	985	13	BX436838	BX436838 BX436838
c	2	53.6	0.8	836	29	CNS04Q18	AL302021 Tetraodon
	3	53.4	0.8	1101	29	CNS003DQ	AL064580 Drosophil
	4	52	0.8	1146	29	CNS021G2	AL176843 Tetraodon
c	5	50.8	0.8	1201	13	BX358198	BX358198 BX358198
c	6	50.4	0.8	997	29	CNS005TE	AL060767 Drosophil
c	7	50.2	0.8	1101	29	CNS0182P	AL108811 Drosophil
c	8	49.4	0.8	460	14	CB472367	CB472367 sn54_C04.
c	9	49	0.8	844	29	AG042177	AG042177 Pan trogl
	10	49	0.8	1101	29	CNS00E7Y	AL069057 Drosophil
c	11	49	0.8	1165	13	BX338369	BX338369 BX338369
c	12	48.8	0.8	1198	13	BX395612	BX395612 BX395612
c	13	48.8	0.8	1201	13	BX385348	BX385348 BX385348
	14	48.2	0.7	440	13	BU497603	BU497603 PfESToab8
	15	48.2	0.7	566	12	BM275442	BM275442 PfESToaa6
	16	48.2	0.7	569	12	BM273626	BM273626 PfESToaa5
c	17	48	0.7	887	13	BX441520	BX441520 BX441520
	18	47.8	0.7	946	13	BX459436	BX459436 BX459436
	19	47.6	0.7	993	13	BQ729165	BQ729165 AGENCOURT
	20	47.4	0.7	1065	9	AL515373	AL515373 AL515373
c	21	47.4	0.7	1101	29	CNS0021J	AL061936 Drosophil
	22	47.4	0.7	1225	29	CNS0161D	AL106171 Drosophil
c	23	47.2	0.7	240	9	AU074526	AU074526 AU074526
c	24	47.2	0.7	944	29	CNS06WVA	AL418940 T3 end of
	25	47.2	0.7	994	13	BX414650	BX414650 BX414650
	26	47.2	0.7	1005	13	BX441452	BX441452 BX441452
c	27	47.2	0.7	1201	9	AL548154	AL548154 AL548154
c	28	47	0.7	712	13	BX416727	BX416727 BX416727
c	29	47	0.7	1201	9	AL558113	AL558113 AL558113
c	30	46.8	0.7	905	29	CNS00KHX	AL077798 Drosophil
c	31	46.8	0.7	913	28	AZ546811	AZ546811 ENTEL88TR
	32	46.8	0.7	975	13	BX414927	BX414927 BX414927
c	33	46.8	0.7	994	29	CNS015XG	AL106030 Drosophil
	34	46.6	0.7	589	28	AQ541309	AQ541309 RPCI-11-3
c	35	46.6	0.7	794	28	BH498125	BH498125 BOGWF71TR
c	36	46.6	0.7	964	29	CNS017VX	AL108567 Drosophil
c	37	46.6	0.7	1101	29	CNS0183Y	AL108856 Drosophil
c	38	46.6	0.7	1161	13	BX332029	BX332029 BX332029
c	39	46.6	0.7	1201	13	BX399692	BX399692 BX399692
	40	46.4	0.7	923	28	AZ538694	AZ538694 ENTEY15TF
	41	46.4	0.7	1101	29	CNS003DX	AL064587 Drosophil
c	42	46.4	0.7	1201	13	BX336719	BX336719 BX336719
c	43	46.4	0.7	1201	29	CNS016AD	AL106495 Drosophil
c	44	46.2	0.7	1101	29	CNS003DQ	AL064580 Drosophil
	45	46.2	0.7	1167	29	CNS07360	AL427102 clone BA0

ALIGNMENTS

RESULT 1
BX436838
LOCUS BX436838 985 bp mRNA linear EST 15-MAY-2003
DEFINITION BX436838 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP003YH18 3-PRIME, mRNA sequence.
ACCESSION BX436838
VERSION BX436838.1 GI:30789527
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 985)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0CAP003DD09NP1.
FEATURES Location/Qualifiers
source 1. .985
/organism="Homo sapiens"

```
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0CAP003YH18"
                     /tissue_type="THYMUS"
                     /clone_lib="Homo sapiens THYMUS"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
BASE COUNT      337 a    120 c     95 g    277 t    156 others
ORIGIN

  Query Match             0.9%;  Score 58;  DB 13;  Length 985;
  Best Local Similarity   46.7%;  Pred. No. 5.5;
  Matches   84;  Conservative   27;  Mismatches   69;  Indels    0;  Gaps    0;

Qy       6303 TAATGCTGTGCAGTGTCACATTAATCGTATATCACACTACATATTAACAACACTATATCA 6362
              ::|:  | |  |   | | ||  ||  ||||| ::: |:  ||||   :: : |:|:|
Db        746 WWAWATTWTAAAAAWTAATATATATATTATATAWHWTTWWTTATTTTTWWTWATWTWTAT 805

Qy       6363 CTTTTATGAGACTCACTATGGGTCTCTAATATACACTACACATATTTTACTTAAAAACAC 6422
               |:|||    : : | |||   | | :|||||  |   |   ||: : ||  :|| |: |
Db        806 ATWTTAATTAWWWWATTATATATATAWAATATAAATAATTTATWAWATATAWAATAWTAT 865

Qy       6423 TATACACACTTTATAAATTCTTTTATAATTTTTCTTTTGTTTTTATTTTGTTTTTAAAAT 6482
              |: : |  ||| ||| | | :|||| ||| | ::|| |:|:|||| | |||| || ||
Db        866 ATTWTWTATATTAAAAAATATAWTATATTTTATWWHTTTWTWTATTATHTTTTAAATAT 925


RESULT 2
CNS04Q18/c
LOCUS       CNS04Q18                836 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            128L13 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL302021
VERSION     AL302021.1  GI:8179132
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 836)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .836
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="128L13"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG128CF07SP1~end :
                     PUC-Ori"
BASE COUNT      301 a     30 c    173 g    299 t     33 others
ORIGIN

  Query Match             0.8%;  Score 53.6;  DB 29;  Length 836;
  Best Local Similarity   57.8%;  Pred. No. 33;
  Matches   89;  Conservative    3;  Mismatches   62;  Indels    0;  Gaps    0;

Qy       6331 ATATCACACTACATATTAACAACACTATATCACTTTTATGAGACTCACTATGGGTCTCTA 6390
              || | |  ||| |||  || :| | || ||   |||||  | |  ||  |  |    |
Db        579 ATYTTAATCTAAATAAAAAAMAAATTAAATTTATTTTAATAAAACCAAATTTTTTTAAAA 520

Qy       6391 ATATACACTACACATATTTTACTTAAAAACACTATACACACTTTATAAATTCTTTTATAA 6450
              |||  :| |  ||||| :|  ||||| | |  || |||| |  | ||  ||||||  |
Db        519 ATACTMATTTTACATAAAWTTATTAAATAAAAMATCCACAATCAAAAATCTCTTTTTAAT 460

Qy       6451 TTTTTCTTTTGTTTTTATTTTGTTTTTAAAATTT 6484
              || ||  ||| ||| |   || ||||| ||||||
Db        459 TTATTTATTTATTTATTCATTTTTTTTCAAATTT 426


RESULT 3
CNS003DQ
LOCUS       CNS003DQ                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08I09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL064580
VERSION     AL064580.1  GI:4941932
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR08I09"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      291 a     51 c    117 g    404 t    238 others
```

```
ORIGIN

  Query Match             0.8%;  Score 53.4;  DB 29;  Length 1101;
  Best Local Similarity   51.5%;  Pred. No. 33;
  Matches   84;  Conservative   14;  Mismatches   65;  Indels    0;  Gaps    0;

Qy       6322 ATTAATCGTATATCACACTACATATTAACAACACTATATCACTTTTATGAGACTCACTAT 6381
              ||:|:|     :: | | || |||| || || |  ||:  |  | ||| | |   |  :|
Db        274 ATWAWTWTATAWWAAWAWTAAATATAAAWAAAAAAATWAWAAATATATAAAAAATATAWT 333

Qy       6382 GGGTCTCTAATATACACTACACATATTTTACTTAAAAACACTATACACACTTTATAAATT 6441
              : |:  |||          |:|||||  |||||| | :|||   | |||  ||| |
Db        334 WWAATWTTWTAATATTTATATATTWTTTTATATAAAAAWAAWATATTTATTTTTAAAAAT 393

Qy       6442 CTTTTATAATTTTTCTTTTGTTTTTATTTTGTTTTTAAAATTT 6484
              ||||   | |||| |||| ||||||:||| || |||:  ::|
Db        394 ATTTTTATAATTTTATTTTWTTTTTAWTTTWTTATTAWTTWWT 436


RESULT 4
CNS021G2
LOCUS       CNS021G2                1146 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            225004 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL176843
VERSION     AL176843.1  GI:7814900
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1146)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1146
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="225004"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG225BH02LP1~end : T7"
BASE COUNT      393 a     60 c     73 g    274 t    346 others
ORIGIN

  Query Match             0.8%;  Score 52;  DB 29;  Length 1146;
  Best Local Similarity   31.7%;  Pred. No. 57;
  Matches  113;  Conservative   71;  Mismatches  172;  Indels    0;  Gaps    0;

Qy       6126 TTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGATGAC 6185
              ||:||  |::: ::::   |: ::  :: |:   ||:: || | |    | : |:|  |
Db        692 TTDTTTAAKKRKKKRKKWAGKGKKRAKKRAKAAKTAKRRTTTTKAKATKAARAARAAAAA 751

Qy       6186 TGAGCGCGGACACTGACATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGTGCAT 6245
                |     ::  :::: ::: :: |||| : ::::|  :   : |:|:: | |  : |
Db        752 AAATTATTKRRKBKRRKKKRRKKKRAAAAKKKBRKKTGRKGRRKGGRARKKGRGAARGAA 811

Qy       6246 AATGCCACGCGCCGCTTGACACTAAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAA 6305
              : :    |       ||:|   |  || | :|: | |||   ::: :   :::  |
Db        812 GRRKKNKKGDAWATTTTDATTATTTAATTTRADATTTTTADARKKKKKKKKKKKKKTTT 871

Qy       6306 TGCTGTGCAGTGTCACATTAATCGTATATCACACTACATATTAACAACACTATATCACTT 6365
              |  | |  : : |     ||  |  | | | | | |  | |||:| ||    | |  |  |
Db        872 TTTTTTTRTKTTTTTTTTTTTTTTTTTTAAATTTTAAATTWAAAATTTAAAAAAAANT 931

Qy       6366 TTATGAGACTCACTATGGGTCTCTAATATACACTACACATATTTTACTTAAAAACACTAT 6425
              |||  | | |    | :       |  :: : : |     ::|||:: | |||||  : |:|
Db        932 TTAAAAAAATTTDAAAKTAAAANTTTKRRDKRTTTTTTTDRTTTWWTTAAAAAAAWTTWT 991

Qy       6426 ACACACTTTATAAATTCTTTTATAATTTTTCTTTTGTTTTTATTTTGTTTTTAAAA 6481
                :  ||  || |      | |:|||     |||| ||||| |::| |||| || :
Db        992 TAWTTTTTATTATAAAWAATAAWAATAAAAATTTTTTTTTTTTTWWTTTTTTAAANR 1047


RESULT 5
BX358198/c
LOCUS       BX358198                1201 bp    mRNA    linear   EST 05-MAY-2003
DEFINITION  BX358198 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI035YE19 5-PRIME, mRNA sequence.
ACCESSION   BX358198
VERSION     BX358198.1  GI:30368196
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 1590.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI035AC10QP1&cluster=1590.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI035AC10QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI035YE19"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      373 a    144 c    192 g    367 t    125 others
ORIGIN

  Query Match             0.8%;  Score 50.8;  DB 13;  Length 1201;
```

```
  Best Local Similarity   39.6%;  Pred. No. 90;
  Matches   89;  Conservative   36;  Mismatches  100;  Indels    0;  Gaps    0;

Qy       6260 CTTGACACTAAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGTC 6319
              | |:::  |||  :|   : | || |   |  |: |::|    | ||| :: |  |   :|
Db       1016 CCTRRSSTTAAYYYTTTTWWTTTTCCAAAAWAMCCMMAAAAAAAAAATTYYTTWAAAAAWC 957

Qy       6320 ACATTAATCGTATATCACACTACATATTAACAACACTATATCACTTTTATGAGACTCACT 6379
              :|    |  | |  |: :| |: | | || |:: |  | |   ||||  | : ::::  |
Db        956 CMAAATTTATTTTTCCMWMCMAMTTTTAAAAAMMCCCCTTTTTTTTTTTTTTWMMYYYYTT 897

Qy       6380 ATGGGTCTCTAATATACACTACACATATTTTACTTAAAAACACTATACACACTTTATAAA 6439
               |:|: |: |  |: | |  |    || :|:  |    :| || ::   : |||  ||:
Db        896 TTKGSMCHTTTTTWAAAAAAATKKKTACCKTWAATTTTNMCCCTTWRGGTWWTTTTAAAR 837

Qy       6440 TTCTTTTATAATTTTTCTTTTGTTTTTATTTTGTTTTTAAAATTT 6484
                :     |  ||||| |||| ||||| |||| ||||| |:|: |
Db        836 GSMCCCCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAWAWAT 792


RESULT 6
CNS005TE/c
LOCUS       CNS005TE                997 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL060767
VERSION     AL060767.1  GI:4943573
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 997)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .997
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR12K22"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT       89 a     99 c     13 g    258 t    538 others
ORIGIN

  Query Match             0.8%;  Score 50.4;  DB 29;  Length 997;
  Best Local Similarity   17.6%;  Pred. No. 1.1e+02;
  Matches   91;  Conservative  169;  Mismatches  258;  Indels    0;  Gaps    0;

Qy       2672 AAGAAGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGAAGAAGCAGCAA 2731
              |:  |:||:|  :: |     :|   ::|:   |   ||   : : ||  : :| :: ||
Db        953 ARAGARAARAGARRAGAGARARAAARRRARRRRAGRRAGRAGRGRSAGGGRGRGGRRAAA 894
```

```
Qy       2732 GCCAAGAGGACGAAACGCAAGCCTAAACCAGGGAAACGACAACGTATGTGTATGAAGTTG 2791
              |::|: : :|:|   ::   :||:    ::::  :: :    : : : | |:|| |
Db        893 ARAARDARARARARARRRARRAKAARARGARRRRGRRRRRGRAGRRRARGAGGRAGRRG 834

Qy       2792 GAGTCGGACAAGACATTTCCGATCATGCTGAACGGCCAAGTGAATGGATATGCCTGCGTT 2851
              : :  ::| |:::        :  : :  :|: ::  | : ::: :||   :   | :
Db        833 RGRGRRRARARRRAGAARARARGRRARAARARARRRRAGRRRRRARGARGRRRRGGRRRG 774

Qy       2852 GTCGGAGGAAGGCTGATGAAACCACTCCACGTTGAAGGAAAAATTGATAATGAGCAATTA 2911
              :  :::|::::   :|  :|:  |    : :  ::|| ||::|  :: :: ::|  | :
Db        773 RGGRRRGRRRRAARRARARARAGARAAARRRRARRAGAAARRARRRRRRRRRGAGARAR 714

Qy       2912 GCGGCCGTGAAATTGAAGAAGGCTAGCATGTACGACTTGGAGTACGGCGACGTTCCCCAG 2971
              : ::  :  |:|  |||::::|   : |   : |:   :::|   :: :: :      |:
Db        713 RARRRARRAARAGAGAARRRRGRRGRGAGARRRGRGRGRRRGAGARRRRRGRRMRTRRAR 654

Qy       2972 AACATGAAATCAGACACGCTGCAGTACACCAGCGACAAACCACCGGGCTTCTACAACTGG 3031
              :: : :: |  |:: | |  : :: :    :: :|  ::  |  |::     : |:  ::
Db        653 RRRRARRGAGAARRRARGARRRRRGRRGRRRRGRARGRRRRAGRGRRRGRRARGARRGRR 594

Qy       3032 CACCACGGCGCAGTCCAGTATGAGAATGGGAGATTTACCGTACCGAGAGGAGTGGGCGGG 3091
              |  : ::   ||   :: : ::|:| :  :||   :  : | ::::  |: | |  |
Db        593 RAGRRRRRARAGAGARRGRRRRGRARRAARGAARRRRARRARRRRRRAAARRGAGRAGA 534

Qy       3092 AAAGGCGACAGCGGAAGACCGATCCTGGACAACAGAGGCAGAGTTGTGGCTATTGTTCTA 3151
              :|:|: |: :: ::|:::  ::    :::  :  |::: |||:  :  :  |  :   |
Db        533 RARGRAGRARRARRARRRRRRRGGARRRGGRRGGRRRAAGARRRRAARAAARARRRARA 474

Qy       3152 GGAGGTGCAAATGAGGGCACGCGTACGGCGCTTTCAGT 3189
              |:||  : :  |                       |   |
Db        473 GRAGARRRRTNTNNNNNNNNNNNNNNNNAGCNTNNNT 436


RESULT 7
CNS0182P/c
LOCUS       CNS0182P                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN37D10 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL108811
VERSION     AL108811.1  GI:5629115
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN37D10"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
BASE COUNT      274 a    268 c    128 g     73 t    358 others
ORIGIN
```

```
  Query Match             0.8%;  Score 50.2;  DB 29;  Length 1101;
  Best Local Similarity   17.6%;  Pred. No. 1.2e+02;
  Matches   39;  Conservative  114;  Mismatches   68;  Indels    0;  Gaps    0;

Qy       6265 CACTAAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGTCACATT 6324
              |:::::::: : :: :::: :  :: ::||  :::::: |  :  : :::|:::
Db       1019 CHHWWWHWMAYAWWWWWHYAWAMHWHHHHMMCAMCMHWYMMHMCCYMCYCCWMHMWCHHH 960

Qy       6325 AATCGTATATCACACTACATATTAACAACACTATATCACTTTTATGAGACTCACTATGGG 6384
              :   ::: |:::::: |:|:||:::: : :::::::| :   |: : | : ::::
Db        959 MTATTWHHYTYHYWMWTCHTMTTWWYHYMTMHWHHHHAAWAMAAHMHAATHTWYHMAHAH 900

Qy       6385 TCTCTAATATACACTACACATATTTTACTTAAAAACACTATACACACTTTATAAATTCTT 6444
              : |::  : | : :: : : : |:|: ::::  ||:::  : ::::::::::||:| ||
Db        899 HTTYYTTYYTYYYYYYYYYYYYTYTYTHHWWTTAAHHHWYAHWHHWHWWYHYWAAWTTTT 840

Qy       6445 TTATAATTTTTCTTTGTTTTTATTTTGTTTTTAAAATTTC 6485
              || |  |||:|::::: ||::| :|::  |:|::   |:|:
Db        839 TTTTTTTTTWTHWHYHHTTYYTTYTYYTCTYTYMCCTTWTY 799


RESULT 8
CB472367/c
LOCUS       CB472367                460 bp    mRNA    linear   EST 26-MAR-2003
DEFINITION  sn54_C04.f sn Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   CB472367
VERSION     CB472367.1  GI:29278753
KEYWORDS    EST.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   Neilan,J.G., Kutish,G.F., Lu,Z., Zsak,A. and Rock,D.L.
  TITLE     Sequence analysis of African swine fever virus infected and
            non-infected porcine macrophage cDNA libraries
  JOURNAL   Unpublished
COMMENT     Contact: Neilan JG
            Plum Island Animal Disease Center
            US Department of Agriculture, Agricultural Research Service
            PO Box 848, Greenport, NY 11944-848, USA
            Tel: 631 323 3133
            Fax: 631 323 3044
            Email: jneilan@piadc.ars.usda.gov
            Single pass sequencing. Bases called with phred v0.020425.c and
            trimmed with the aid of the trim_alt option. Vector identified by
            cross_match v0.990329 and lucy v1.17p.
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .460
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /tissue_type="lymphoid"
                     /cell_type="macrophage"
                     /lab_host="DH10B"
                     /clone_lib="sn"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI;
                     Library made from pools of polyA selected RNA, isolated at
                     different times post-infection (0 to 16 hrs) from African
                     swine fever virus (isolate Pretorisuskop/96/4) infected
                     swine macrophages. Macrophages were derived from
                     peripheral blood mononuclear cells cultured for 48 hrs on
                     plastic in the presence of 30% L929 supernatant."
BASE COUNT      264 a     22 c     91 g     83 t
ORIGIN

  Query Match             0.8%;  Score 49.4;  DB 14;  Length 460;
  Best Local Similarity   55.6%;  Pred. No. 2e+02;
  Matches   95;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy       6314 AGTGTCACATTAATCGTATATCACACTACATATTAACAACACTATATCACTTTTATGAGA 6373
              | | |  | |||||  || ||     |   | ||   || | | | | || ||| |  ||
Db        196 ATTTTTTCTTTAATTTTAAATTTTTTTTTTCTTTTTAAAAGTTTTTTACCTTTTTTGGA 137

Qy       6374 CTCACTATGGGTCTCTAATATACACTACACATATTTTACTTAAAAACACTATACACACTT 6433
              | |   | | |  |  |     |     | ||||  ||||||||  | |       ||
Db        136 TATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAAAACTTTTTTTTGTTT 77

Qy       6434 TATAAATTCTTTTATAATTTTTCTTTTGTTTTTATTTTGTTTTTAAAATTT 6484
              | |    || |||||   ||||| |||| ||||| |||| ||||||||||||
Db         76 TTTTTTTTTTTTTACCTTTTTTTTTTTTTTTTTTTTTTTTTTAAAATTT 26


RESULT 9
AG042177/c
LOCUS       AG042177                844 bp    DNA     linear   GSS 01-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-020B19.F, genomic survey sequence.
ACCESSION   AG042177
VERSION     AG042177.1  GI:16570902
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 844)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .844
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-020B19.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      243 a     43 c     66 g    487 t      5 others
ORIGIN

  Query Match             0.8%;  Score 49;  DB 29;  Length 844;
  Best Local Similarity   51.4%;  Pred. No. 2e+02;
  Matches  112;  Conservative    0;  Mismatches  106;  Indels    0;  Gaps    0;

Qy       6264 ACACTAAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGTCACAT 6323
              | |  |||| |       || |   |  ||| |   | || ||        |  ||    |
Db        601 ATAAGAAAATTAATATAATATATAATAAAGAATTATACAAAAAATAAAAAAAAGTNTATT 542

Qy       6324 TAATCGTATATCACACTACATATTAACAACACTATATCACTTTTATGAGACTCACTATGG 6383
              |||  || |  | | ||     ||| || |  |||| |  ||||| | |   | |||
Db        541 AAATAATAAAAAATAATAATATATAAAAAAATAATATAATATTTATTAAAAAAAATATAA 482

Qy       6384 GTCTCTAATATACACTACACATATTTTACTTAAAAACACTATACACACTTTATAAATTCT 6443
              | |  || | | |  | |||| |    || ||  | |||| | | |  ||||||
Db        481 ATATAAAAAAGATAAAAATAATATATACAATAGAATAAATATATAAAATAAATAAATATA 422
```

```
Qy       6444 TTTATAATTTTTCTTTTGTTTTTATTTTGTTTTTAAAA 6481
              || | ||||     ||| |||| ||||| |||  | ||
Db        421 TTAAAAATTAGGTATTTTTTTTAATTTTTTTTAAATAA 384


RESULT 10
CNS00E7Y
LOCUS       CNS00E7Y                1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR28A16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069057
VERSION     AL069057.1  GI:4949017
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR28A16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      248 a    175 c    173 g    341 t    164 others
ORIGIN

  Query Match             0.8%;  Score 49;  DB 29;  Length 1101;
  Best Local Similarity   38.1%;  Pred. No. 1.9e+02;
  Matches   61;  Conservative   44;  Mismatches   55;  Indels    0;  Gaps    0;

Qy       6318 TCACATTAATCGTATATCACACTACATATTAACAACACTATATCACTTTTATGAGACTCA 6377
              ||:::::| :| :|: :|: : ::   :||  ::|   | |:|:: |||| | |   | |
Db        910 TCMMMHHAYHCMHAHTHCHTMWYWWTAWTTTTMMAWTTTTTWTMHWTTTTTTTATTTTWA 969

Qy       6378 CTATGGGTCTCTAATATACACTACACATATTTTACTTAAAAACACTATACACACTTTATA 6437
                :|   ::|: || |||   |   :|:| | |: || ::|: | |: ::: : | :
Db        970 AMHTTHYWMTMMAAAATAWTTTTATHAWAATMTMTTTTWWAWWAWTWAWHWWHTTCWYNY 1029

Qy       6438 AATTCTTTTATAATTTTTCTTTTGTTTTTATTTTGTTTTT 6477
              :|| ||||:| ::| :|:|| : ||||: |||| ||:||
Db       1030 CMTTTTTTTWTTMHTMHTHTTMWTTTTTWTTTTTTTTYTT 1069


RESULT 11
BX338369/c
LOCUS       BX338369                1165 bp    mRNA    linear   EST 02-MAY-2003
DEFINITION  BX338369 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI058YF14 5-PRIME, mRNA sequence.
ACCESSION   BX338369
VERSION     BX338369.1  GI:30343682
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1165)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI058DC07QP1.
FEATURES             Location/Qualifiers
     source          1. .1165
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI058YF14"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      567 a    109 c    140 g    251 t     98 others
ORIGIN

  Query Match             0.8%;  Score 49;  DB 13;  Length 1165;
  Best Local Similarity   41.1%;  Pred. No. 1.8e+02;
  Matches   67;  Conservative   33;  Mismatches   63;  Indels    0;  Gaps    0;

Qy       6322 ATTAATCGTATATCACACTACATATTAACAACACTATATCACTTTTATGAGACTCACTAT 6381
              ::: :|   : |  | : :| ::| ||: |: | | :|| |    |   : |   | :||
Db        977 WWWTWTWWAWAAAWATWAWATWWAATAWWAWTAWTTWATWAAAAATTAAWTATAAAAWAT 918

Qy       6382 GGGTCTCTAATATACACTACACATATTTTACTTAAAAACACTATACACACTTTATAAATT 6441
                |   || |:|:    | : :|:||||| :|||| | | |:|| | |    : |:|||
Db        917 WTTTTATTATTWTWTTTAATWTWTWTTTTATWTAAATATATTWTAWAAAAAAAWAAWATT 858

Qy       6442 CTTTTATAATTTTTCTTTTGTTTTTATTTTGTTTTTAAAATTT 6484
              :::||: :|| ::|: ||  |:||| |||: |||||    |||.
Db        857 YYWTTWAWATAWYTMATTAATWTTTTTTTTWWTTTTTTTTTTTT 815


RESULT 12
BX395612/c
LOCUS       BX395612                1198 bp    mRNA    linear   EST 13-MAY-2003
DEFINITION  BX395612 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI002YP22 3-PRIME, mRNA sequence.
ACCESSION   BX395612
VERSION     BX395612.1  GI:30616740
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1198)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
```

```
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI002DH11NP1.
FEATURES              Location/Qualifiers
     source           1. .1198
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CS0DI002YP22"
                      /tissue_type="PLACENTA COT 25-NORMALIZED"
                      /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                      /note="1st strand cDNA was primed with a NotI-oligo(dT)
                      primer. Five prime end enriched, double-strand cDNA was
                      digested with Not I and cloned into the Not I and EcoR V
                      sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      242 a    122 c    196 g    372 t    266 others
ORIGIN

  Query Match             0.8%;  Score 48.8;  DB 13;  Length 1198;
  Best Local Similarity   35.6%;  Pred. No. 2e+02;
  Matches   74;  Conservative   47;  Mismatches   87;  Indels    0;  Gaps    0;

Qy       6278 TGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGTCACATTAATCGTATATCAC 6337
             |:: ::|: ::  ||  | | :  |:  |: | |  :     :::  ||   |:|:|| :
Db        952 TDKTKKTMDKWAAAAAKAAAAYAAAKTKTKTTTTMAMCATCTMMMAMAAATTTWTWTCCM 893

Qy       6338 ACTACATATTAACAACACTATATCACTTTTATGAGACTCACTATGGGTCTCTAATATACA 6397
             | ||    : | |: :  |: | :: ||:||: |   |:::| |   : | ::| | : :
Db        892 AWTAWKMTWAACCMCMTTTMMAAMMATTWTAYMAATTTMMMTTTTTAHTTAKMACAAMAM 833

Qy       6398 CTACACATATTTTACTTAAAAACACTATACACACTTTATAAATTCTTTTATAATTTTTCT 6457
             : |:  :|||| |  |||| ||   :|:: |  :||:: |||        | |||:|
Db        832 MAAMTTWTATTATTTTTAAKAATTTKAKMAATKYTTKMMAAAAMWMMMATTKATTWTMAA 773

Qy       6458 TTTGTTTTTATTTTGTTTTTAAAATTTC 6485
             ||  |||   |:||  |||| :|:||:|
Db        772 TTCMTTTCATTWTTTCTTTTCHAWTTKC 745


RESULT 13
BX385348/c
LOCUS       BX385348               1201 bp    mRNA    linear   EST 08-MAY-2003
DEFINITION  BX385348 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
            Homo sapiens cDNA clone CS0DL010YM11 5-PRIME, mRNA sequence.
ACCESSION   BX385348
VERSION     BX385348.1  GI:30451334
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 727.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DL010AG06QP1&cluster=727.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DL010AG06QP1.
FEATURES              Location/Qualifiers
     source           1. .1201
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CS0DL010YM11"
                      /cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
                      /cell_line="RAMOS CELL LINE"
                      /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
                      25-NORMALIZED"
                      /note="1st strand cDNA was primed with a NotI-oligo(dT)
                      primer. Five prime end enriched, double-strand cDNA was
                      digested with Not I and cloned into the Not I and EcoR V
                      sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      473 a     50 c    184 g    338 t    156 others
ORIGIN

  Query Match             0.8%;  Score 48.8;  DB 13;  Length 1201;
  Best Local Similarity   56.6%;  Pred. No. 2e+02;
  Matches   86;  Conservative    2;  Mismatches   64;  Indels    0;  Gaps    0;

Qy       6328 CGTATATCACACTACATATTAACAACACTATATCACTTTTATGAGACTCACTATGGGTCT 6387
             |: | | || | || | ||  | || |  ||||| | || |   |       |       |
Db        255 CKAAAAACATATTAAAAATACAAAAAAACATATCATTATTCTCTTAAATCACAACCAAAT 196

Qy       6388 CTAATATACACTACACATATTTTACTTAAAAACACTATACACACTTTATAAATTCTTTTA 6447
             | || ||| | |  |||| ||| || ||||    ||:||   | ||| |   |  ||||
Db        195 CCAAAATAAAATKTACATTTTTCACCTAAACTACCTWTATTTATTTTTTTTTTATTTTTT 136

Qy       6448 TAATTTTTCTTTTGTTTTTATTTTGTTTTTAA 6479
             |  ||||  |||| |||| |||||   |||||
Db        135 TCTTTTTCTTTTTTTTTTCATTTTCCATTTAA 104


RESULT 14
BU497603
LOCUS       BU497603               440 bp    mRNA    linear   EST 11-SEP-2002
DEFINITION  PfESToab81g06.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium
            falciparum 3D7 cDNA 5', mRNA sequence.
ACCESSION   BU497603
VERSION     BU497603.1  GI:22793797
KEYWORDS    EST.
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 440)
  AUTHORS   Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,
            Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B.,
            Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,
            Tsagareishvili,R., Belaygorod,L., Franklin,C., Carr,L.; Grow,A.,
            Maguire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D.,
            Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU Plasmodium EST Project
  JOURNAL   Unpublished
COMMENT     Contact: L. David Sibley
            WashU Plasmodium EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library was constructed by Debopam Chakrabarti DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: L. David Sibley
            (sibley@borcim.wustl.edu)., Washington University
            Seq primer: -40UP from Gibco.
FEATURES              Location/Qualifiers
     source           1. .440
                      /organism="Plasmodium falciparum 3D7"
                      /mol_type="mRNA"
                      /db_xref="taxon:36329"
                      /lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
                      /clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
                      /note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
                      XhoI; Library was constructed by Debopam Chakrabarti.
                      Total RNA samples were isolated from mixed stage
                      saponin(0.1%)-lysed P. falciparum 3D7 infected
                      erythrocytes by the acidic guanidinium-phenol chloroform
                      method. The poly A+ RNA was isolated by the polyAT-Tract
```

mRNA isolation system (Promega, WI) using streptavidin Magnesphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of l ZapII vector using the Zap cDNA synthesis kit (Stragene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the ExAssist helper phage (Stratagene), the phagemids were precitptated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."

BASE COUNT 195 a 22 c 14 g 208 t 1 others
ORIGIN

Query Match 0.7%; Score 48.2; DB 13; Length 440;
Best Local Similarity 55.8%; Pred. No. 3.3e+02;
Matches 92; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```
Qy       6320 ACATTAATCGTATATCACACTACATATTAACAACACTATATCACTTTTATGAGACTCACT 6379
              | || |    |||||     || ||||| | || |  | |     |   |  |    |
Db        215 ATATAATATATATATTTTTATATATATTTAAAAAAAAAAAAAATATAAAAATATTTTATAT 274

Qy       6380 ATGGGTCTCTAATATACACTACACATATTTTACTTAAAAACACTATACACACTTTATAAA 6439
              ||   | | |||||||   || | |||| |   |  |||| |  ||| | |  | ||||
Db        275 ATTAATATATAATATATTATATATATATATATTTATAAAAAAAAAATAAATAAATAATAAT 334

Qy       6440 TTCTTTTATAATTTTTCTTTTGTTTTTATTTTGTTTTTAAAATTT 6484
              | | ||||| ||| |  | ||| | ||  || ||||| || |||
Db        335 ATATATTATATTTTCTTATATGTATATAAATTTTTTTTTAATTTT 379
```

RESULT 15
BM275442
LOCUS BM275442 566 bp mRNA linear EST 20-DEC-2001
DEFINITION PfESToaa68h12.y1 Plasmodium falciparum 3D7 gametocyte cDNA library Plasmodium falciparum 3D7 cDNA 5' similar to SW:UBC1_DROME P25867 UBIQUITIN-CONJUGATING ENZYME E2-17 KD ;, mRNA sequence.
ACCESSION BM275442
VERSION BM275442.1 GI:17968772
KEYWORDS EST.
SOURCE Plasmodium falciparum 3D7
ORGANISM Plasmodium falciparum 3D7
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 566)
AUTHORS Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D., Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I., Tsagareishvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A., Maguire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D., Waterston,R., Wilson,R. and Sibley,D.
TITLE WashU Plasmodium EST Project
JOURNAL Unpublished
COMMENT Contact: L. David Sibley
WashU Plasmodium EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library was constructed by R. Haywood. DNA sequencing by:
Washington University Genome Sequencing Center For information on obtaining a clone please contact: L. David Sibley (sibley@borcim.wustl.edu), Washington University
Seq primer: -40UP from Gibco
High quality sequence stop: 423.
FEATURES Location/Qualifiers
source 1. .566
/organism="Plasmodium falciparum 3D7"
/mol_type="mRNA"
/db_xref="taxon:36329"
/dev_stage="gametocyte (stage III-V)"
/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
/clone_lib="Plasmodium falciparum 3D7 gametocyte cDNA library"
/note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2: XhoI; The library was constructed by R Haywood. cDNAs were synthesized from ganetocyte poly(A)+ RNA by oligo d(T) priming, size-selected and directionally cloned into the EcoRi (5' end) to XhoI (3' end) sites of the Uni-ZAP XR lambda vector (Stratagene). The primary library was mass excised as phagemid using the ExAssist helper phage (Stratagene), the phagemids were precitptated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells. Clone Availability: David Sibley, Washington University."

BASE COUNT 256 a 63 c 53 g 194 t
ORIGIN

Query Match 0.7%; Score 48.2; DB 12; Length 566;
Best Local Similarity 57.7%; Pred. No. 3.1e+02;
Matches 86; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy       6336 ACACTACATATTAACAACACTATATCACTTTTATGAGACTCACTATGGGTCTCTAATATA 6395
              | | || ||||| | |  | |||||     |  || | |    | |||    | | |||||||
Db          1 AGAATATATATTTATATAAATATATATTATGAATTAAATGAATTATATATATATAATATA 60

Qy       6396 CACTACACATATTTTACTTAAAAACACTATACACACTTTATAAATTCTTTTATAATTTTT 6455
              | || |  ||| |   ||| ||||| |||| | |  | |  | |   | ||| ||  |
Db         61 TAATATATTTATATATTTTATAAACAATATAAAAATATAAATATTATATATATTATAATA 120

Qy       6456 CTTTTGTTTTTATTTTGTTTTTAAAATTT 6484
              | | |  || |||| ||||| | | ||
Db        121 TATATATACTTCTTTTTTTTTTCACACTT 149
```

Search completed: November 16, 2003, 08:59:44
Job time : 12079 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 17:26:43 ; Search time 1496 Seconds
(without alignments)
10923.143 Million cell updates/sec

Title: US-10-023-649-1_COPY_1_5000
Perfect score: 5000
Sequence: 1 accctacaaactaatcgatc.....ataattctagaaacgccacc 5000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5000	100.0	11484	12	US-10-023-649-1	Sequence 1, Appli
2	2116.4	42.3	12523	10	US-09-991-258-1	Sequence 1, Appli
3	2113.2	42.3	12379	10	US-09-991-258-14	Sequence 14, Appl
4	2113.2	42.3	13584	10	US-09-991-258-17	Sequence 17, Appl
5	2111	42.2	7479	10	US-09-991-258-2	Sequence 2, Appli
6	1567.2	31.3	10610	11	US-09-994-412-3	Sequence 3, Appli
7	1564	31.3	11517	10	US-09-901-106-1	Sequence 1, Appli
8	1564	31.3	15538	12	US-09-190-246-1	Sequence 1, Appli
9	1473.8	29.5	8100	12	US-09-190-246-4	Sequence 4, Appli
10	1300.4	26.0	8000	11	US-09-507-362-102	Sequence 102, App
11	1300.4	26.0	11740	11	US-09-507-362-103	Sequence 103, App
12	1295.6	25.9	8000	11	US-09-507-362-101	Sequence 101, App
13	1295.6	25.9	11282	10	US-09-733-042-1	Sequence 1, Appli
14	1295.6	25.9	11282	11	US-09-275-883-1	Sequence 1, Appli
15	330.8	6.6	705	12	US-10-123-101-51	Sequence 51, Appl
16	211.4	4.2	5076	10	US-09-991-258-8	Sequence 8, Appli
17	211.4	4.2	6989	10	US-09-991-258-11	Sequence 11, Appl
18	90.8	1.8	282	12	US-09-190-246-7	Sequence 7, Appli
19	50.6	1.0	170	12	US-10-123-101-90	Sequence 90, Appl
20	49	1.0	327	12	US-10-123-101-65	Sequence 65, Appl
21	49	1.0	346	12	US-10-123-101-67	Sequence 67, Appl
22	48.6	1.0	132	12	US-10-123-101-68	Sequence 68, Appl
23	40	0.8	416	12	US-10-027-632-80186	Sequence 80186, A
24	40	0.8	416	12	US-10-027-632-80187	Sequence 80187, A
25	40	0.8	416	12	US-10-027-632-315015	Sequence 315015,
26	40	0.8	416	12	US-10-027-632-315016	Sequence 315016,
27	40	0.8	416	13	US-10-027-632-80186	Sequence 80186, A
28	40	0.8	416	13	US-10-027-632-80187	Sequence 80187, A
29	40	0.8	416	13	US-10-027-632-315015	Sequence 315015,
30	40	0.8	416	13	US-10-027-632-315016	Sequence 315016,
31	39.6	0.8	4610	10	US-09-884-001-1	Sequence 1, Appli
32	39.6	0.8	6479	13	US-10-098-841-101	Sequence 101, App
33	39.2	0.8	775	12	US-10-140-472-120	Sequence 120, App
34	39.2	0.8	775	12	US-10-141-761-120	Sequence 120, App
35	39.2	0.8	775	12	US-10-142-885-120	Sequence 120, App
36	39.2	0.8	775	12	US-10-158-790-120	Sequence 120, App
37	39.2	0.8	775	12	US-10-137-871-120	Sequence 120, App
38	39.2	0.8	775	12	US-10-140-805-120	Sequence 120, App
39	39.2	0.8	775	12	US-10-140-864-120	Sequence 120, App
40	39.2	0.8	775	12	US-10-140-923-120	Sequence 120, App
41	39.2	0.8	775	12	US-10-141-756-120	Sequence 120, App
42	39.2	0.8	775	12	US-10-141-759-120	Sequence 120, App
43	39.2	0.8	775	14	US-10-123-155-120	Sequence 120, App
44	39.2	0.8	775	15	US-10-146-731-120	Sequence 120, App
45	39.2	0.8	1184	12	US-10-140-472-412	Sequence 412, App

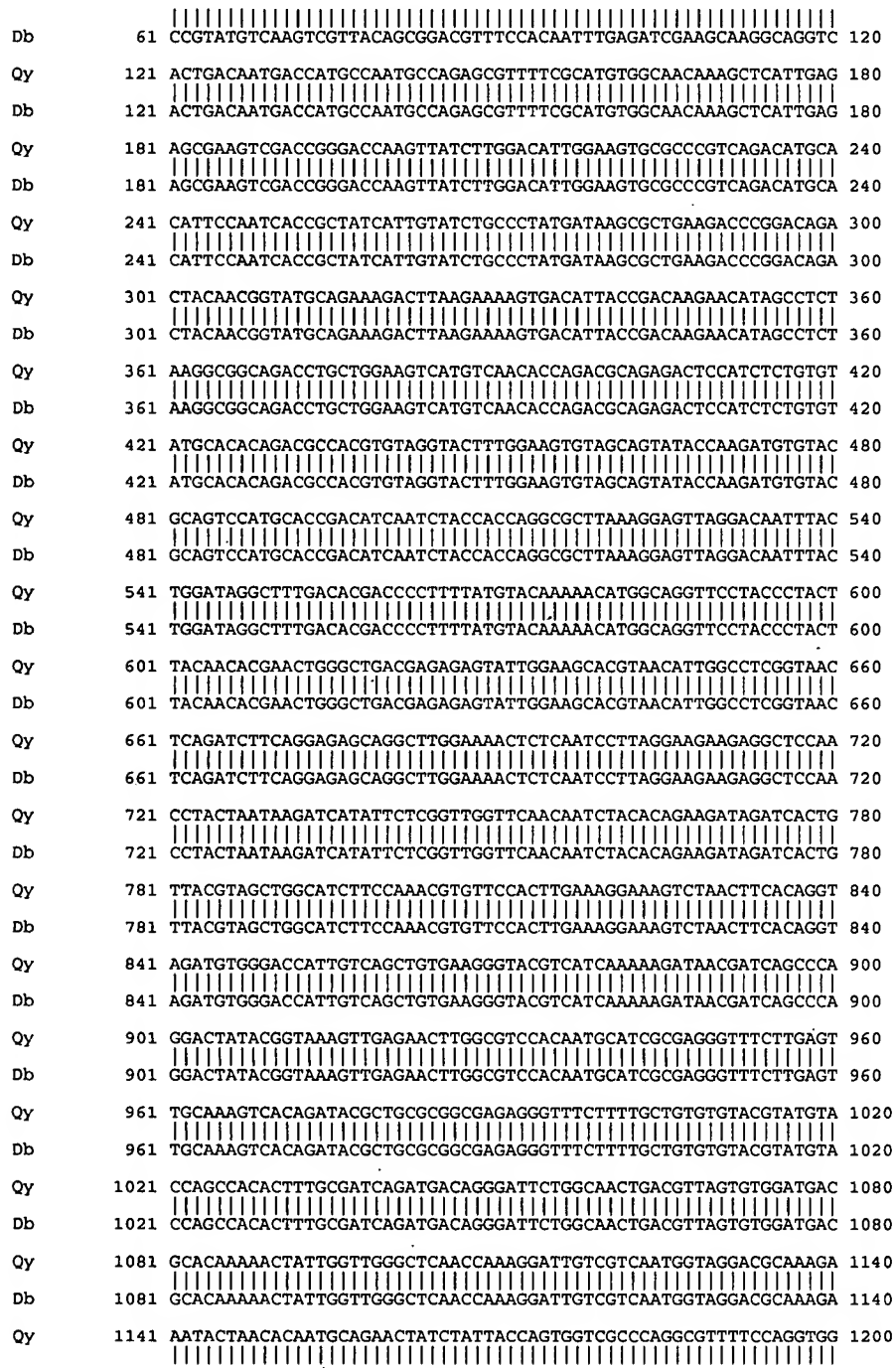
ALIGNMENTS

RESULT 1
US-10-023-649-1
; Sequence 1, Application US/10023649
; Publication No. US20030143201A1
; GENERAL INFORMATION:
; APPLICANT: The Minister of National Defence, Government of Canada
; APPLICANT: Nagata, Leslie P
; APPLICANT: Wong, Jonathon P
; TITLE OF INVENTION: No. US20030143201A1el DNA-Based Vaccine Against the Encephalitis
; FILE REFERENCE: NEL-001
; CURRENT APPLICATION NUMBER: US/10/023,649
; CURRENT FILING DATE: 2002-12-27
; PRIOR APPLICATION NUMBER: 60/256,948
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 11484
; TYPE: DNA
; ORGANISM: Western equine encephalomyelitis virus - strain 71V-1658
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (25)..(7428)
; OTHER INFORMATION: 5' UTR <1 .. 24
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7473)..(11183)
; OTHER INFORMATION: encodes nucleocapsid, E3, E2, 6K and E1 proteins
US-10-023-649-1

Query Match 100.0%; Score 5000; DB 12; Length 11484;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACCCTACAAACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGC 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 ACCCTACAAACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGC 60

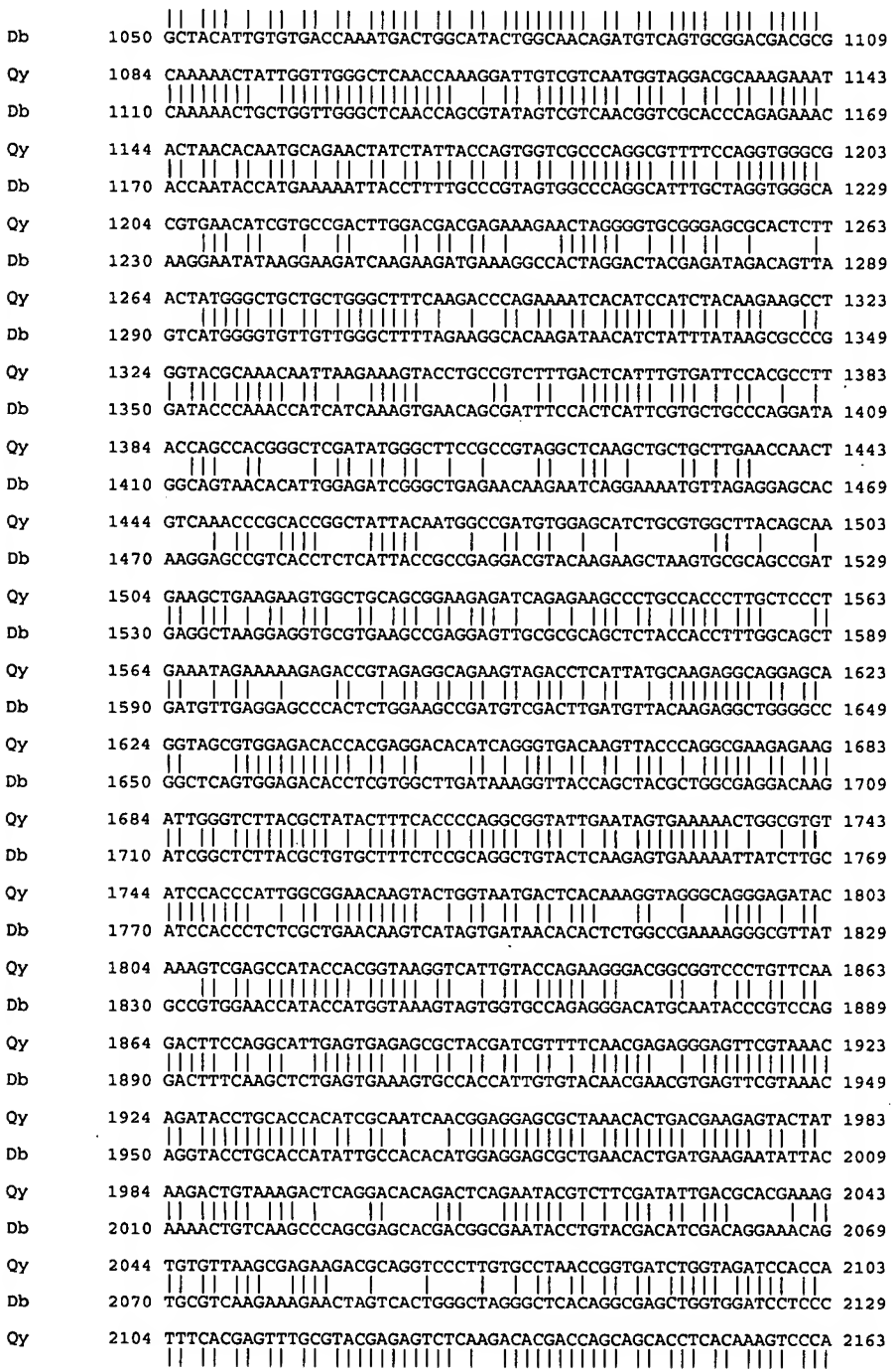
Qy 61 CCGTATGTCAAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTC 120

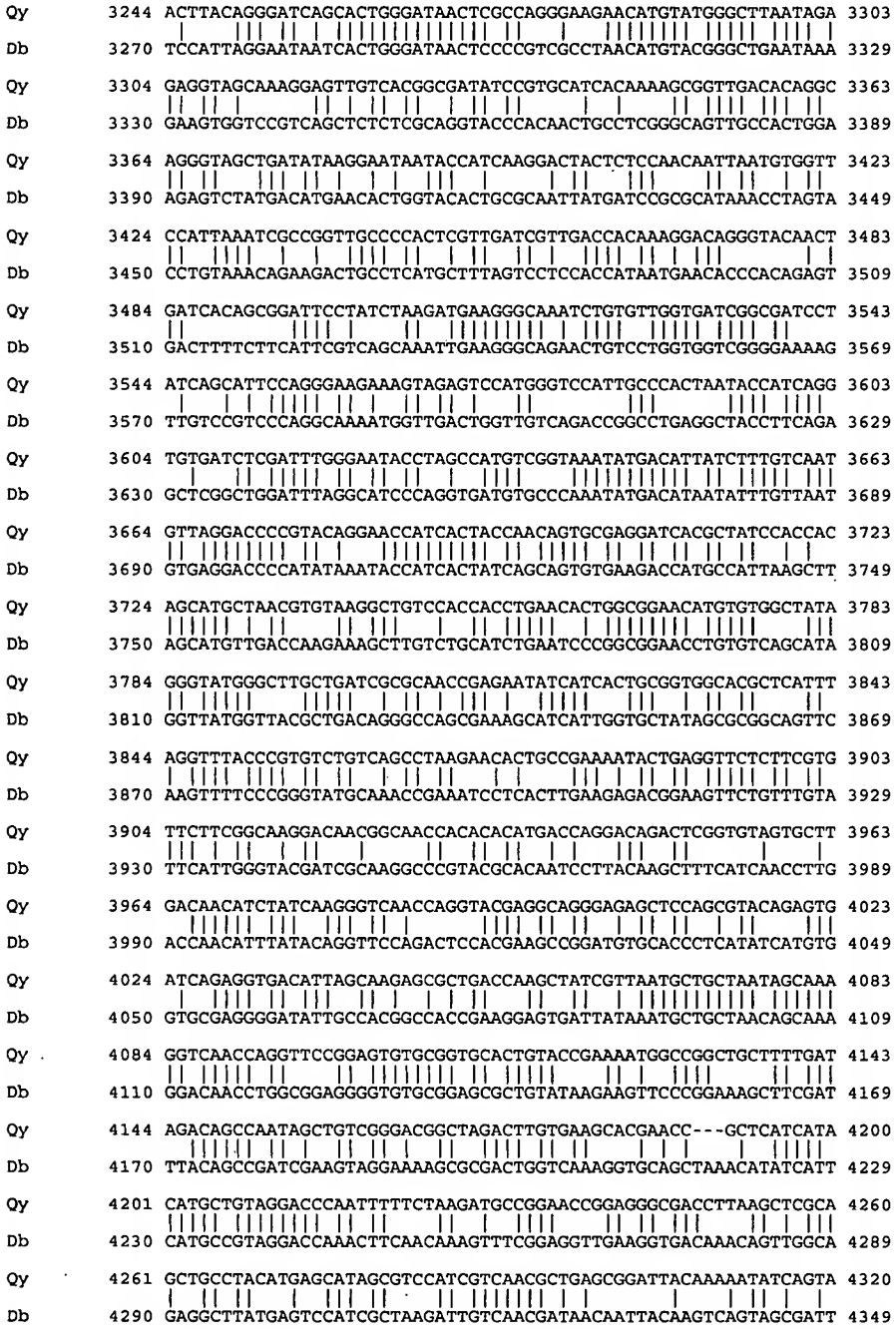

Db 61 CCGTATGTCAAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTC 120
Qy 121 ACTGACAATGACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAG 180
Db 121 ACTGACAATGACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAG 180
Qy 181 AGCGAAGTCGACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCA 240
Db 181 AGCGAAGTCGACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCA 240
Qy 241 CATTCCAATCACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGA 300
Db 241 CATTCCAATCACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGA 300
Qy 301 CTACAACGGTATGCAGAAAGACTTAAGAAAAGTGACATTACCGACAAGAACATAGCCTCT 360
Db 301 CTACAACGGTATGCAGAAAGACTTAAGAAAAGTGACATTACCGACAAGAACATAGCCTCT 360
Qy 361 AAGGCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGT 420
Db 361 AAGGCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGT 420
Qy 421 ATGCACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTAC 480
Db 421 ATGCACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTAC 480
Qy 481 GCAGTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTAC 540
Db 481 GCAGTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTAC 540
Qy 541 TGGATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACT 600
Db 541 TGGATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACT 600
Qy 601 TACAACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAAC 660
Db 601 TACAACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAAC 660
Qy 661 TCAGATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAA 720
Db 661 TCAGATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAA 720
Qy 721 CCTACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTG 780
Db 721 CCTACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTG 780
Qy 781 TTACGTAGCTGGCATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGT 840
Db 781 TTACGTAGCTGGCATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGT 840
Qy 841 AGATGTGGGACCATTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCA 900
Db 841 AGATGTGGGACCATTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCA 900
Qy 901 GGACTATACGGTAAAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGT 960
Db 901 GGACTATACGGTAAAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGT 960
Qy 961 TGCAAAGTCACAGATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTA 1020
Db 961 TGCAAAGTCACAGATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTA 1020
Qy 1021 CCAGCCACACTTTGCGATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGAC 1080
Db 1021 CCAGCCACACTTTGCGATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGAC 1080
Qy 1081 GCACAAAAACTATTGGTTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGA 1140
Db 1081 GCACAAAAACTATTGGTTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGA 1140
Qy 1141 AATACTAACACAATGCAGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGG 1200
Db 1141 AATACTAACACAATGCAGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGG 1200
Qy 1201 GCGCGTGAACATCGTGCCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACT 1260
Db 1201 GCGCGTGAACATCGTGCCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACT 1260
Qy 1261 CTTACTATGGGCTGCTGCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAGAAG 1320
Db 1261 CTTACTATGGGCTGCTGCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAGAAG 1320
Qy 1321 CCTGGTACGCAAACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGC 1380
Db 1321 CCTGGTACGCAAACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGC 1380
Qy 1381 CTTACCAGCCACGGGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCA 1440
Db 1381 CTTACCAGCCACGGGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCA 1440
Qy 1441 ACTGTCAAACCCGCACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAG 1500
Db 1441 ACTGTCAAACCCGCACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAG 1500
Qy 1501 CAAGAAGCTGAAGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTC 1560
Db 1501 CAAGAAGCTGAAGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTC 1560
Qy 1561 CCTGAAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGA 1620
Db 1561 CCTGAAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGA 1620
Qy 1621 GCAGGTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAG 1680
Db 1621 GCAGGTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAG 1680
Qy 1681 AAGATTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCG 1740
Db 1681 AAGATTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCG 1740
Qy 1741 TGTATCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGA 1800
Db 1741 TGTATCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGA 1800
Qy 1801 TACAAAGTCGAGCCATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTT 1860
Db 1801 TACAAAGTCGAGCCATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTT 1860
Qy 1861 CAAGACTTCCAGGCATTGAGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTA 1920
Db 1861 CAAGACTTCCAGGCATTGAGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTA 1920
Qy 1921 AACAGATACCTGCACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTAC 1980
Db 1921 AACAGATACCTGCACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTAC 1980
Qy 1981 TATAAGACTGTAAAGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGA 2040
Db 1981 TATAAGACTGTAAAGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGA 2040
Qy 2041 AAGTGTGTTAAGCGAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCA 2100
Db 2041 AAGTGTGTTAAGCGAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCA 2100
Qy 2101 CCATTTCACGAGTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTC 2160
Db 2101 CCATTTCACGAGTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTC 2160
Qy 2161 CCAACCATCGGAGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCT 2220
Db 2161 CCAACCATCGGAGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCT 2220
Qy 2221 GTGACTAAGAAAGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGG 2280
Db 2221 GTGACTAAGAAAGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGG 2280

```
Qy      2281 GATGTAAGGAGGATGAGACGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTA 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2281 GATGTAAGGAGGATGAGACGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTA 2340

Qy      2341 AATGGGGTTAAGCACCCCGTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCA 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2341 AATGGGGTTAAGCACCCCGTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCA 2400

Qy      2401 GGGACGCTGCTGGCACTGATTGCCATCGTCAAACCTAAGAAAGTGGTATTGTGCGGGGAC 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2401 GGGACGCTGCTGGCACTGATTGCCATCGTCAAACCTAAGAAAGTGGTATTGTGCGGGGAC 2460

Qy      2461 CCAAAACAATGCGGCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATA 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2461 CCAAAACAATGCGGCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATA 2520

Qy      2521 TGCACTGAAGTGTACCATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATC 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2521 TGCACTGAAGTGTACCATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATC 2580

Qy      2581 GTCTCCACGCTCTTCTACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATC 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2581 GTCTCCACGCTCTTCTACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATC 2640

Qy      2641 ATCATAGATACCACAGGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTC 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2641 ATCATAGATACCACAGGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTC 2700

Qy      2701 AGAGGATGGGTGAAACAGCTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCT 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2701 AGAGGATGGGTGAAACAGCTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCT 2760

Qy      2761 GCATCGCAAGGACTTACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAAT 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2761 GCATCGCAAGGACTTACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAAT 2820

Qy      2821 CCACTCTACTCGCAGACTTCTGAGCACGTGAACGTGTTACTTACACGCACAGAAAAACGC 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2821 CCACTCTACTCGCAGACTTCTGAGCACGTGAACGTGTTACTTACACGCACAGAAAAACGC 2880

Qy      2881 ATTGTCTGGAAGACGCTAGCTGGTGATCCCTGGATAAAGACACTTACAGCTAAATATCCC 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2881 ATTGTCTGGAAGACGCTAGCTGGTGATCCCTGGATAAAGACACTTACAGCTAAATATCCC 2940

Qy      2941 GGGGATTTCACGGCTTCATTGGACGACTGGCAGCGCGAACACGACGCCATTATGGCACGC 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2941 GGGGATTTCACGGCTTCATTGGACGACTGGCAGCGCGAACACGACGCCATTATGGCACGC 3000

Qy      3001 GTTCTTGATAAGCCGCAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCG 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3001 GTTCTTGATAAGCCGCAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCG 3060

Qy      3061 AAGGCTTTAGAGCCAGTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAG 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3061 AAGGCTTTAGAGCCAGTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAG 3120

Qy      3121 ACGTTGCACCCATTCAAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTT 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3121 ACGTTGCACCCATTCAAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTT 3180

Qy      3181 TGCACCAGGTTCTTTGGAGTAGACCTGGACAGTGGGTTATTTTCCGCTCCTACCGTCGCA 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3181 TGCACCAGGTTCTTTGGAGTAGACCTGGACAGTGGGTTATTTTCCGCTCCTACCGTCGCA 3240

Qy      3241 CTTACTTACAGGGATCAGCACTGGGATAACTCGCCAGGGAAGAACATGTATGGGCTTAAT 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3241 CTTACTTACAGGGATCAGCACTGGGATAACTCGCCAGGGAAGAACATGTATGGGCTTAAT 3300

Qy      3301 AGAGAGGTAGCAAAGGAGTTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACA 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3301 AGAGAGGTAGCAAAGGAGTTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACA 3360

Qy      3361 GGCAGGGTAGCTGATATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTG 3420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3361 GGCAGGGTAGCTGATATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTG 3420

Qy      3421 GTTCCATTAAATCGCCGGTTGCCCCACTCGTTGATCGTTGACCACAAAGGACAGGGTACA 3480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3421 GTTCCATTAAATCGCCGGTTGCCCCACTCGTTGATCGTTGACCACAAAGGACAGGGTACA 3480

Qy      3481 ACTGATCACAGCGGATTCCTATCTAAGATGAAGGGCAAATCTGTGTTGGTGATCGGCGAT 3540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3481 ACTGATCACAGCGGATTCCTATCTAAGATGAAGGGCAAATCTGTGTTGGTGATCGGCGAT 3540

Qy      3541 CCTATCAGCATTCCAGGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATACCATC 3600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3541 CCTATCAGCATTCCAGGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATACCATC 3600

Qy      3601 AGGTGTGATCTCGATTTGGGAATACCTAGCCATGTCGGTAAATATGACATTATCTTTGTC 3660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3601 AGGTGTGATCTCGATTTGGGAATACCTAGCCATGTCGGTAAATATGACATTATCTTTGTC 3660

Qy      3661 AATGTTAGGACCCCGTACAGGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCAC 3720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3661 AATGTTAGGACCCCGTACAGGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCAC 3720

Qy      3721 CACAGCATGCTAACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCT 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3721 CACAGCATGCTAACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCT 3780

Qy      3781 ATAGGGTATGGGCTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCACGCTCA 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3781 ATAGGGTATGGGCTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCACGCTCA 3840

Qy      3841 TTTAGGTTTACCCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTC 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3841 TTTAGGTTTACCCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTC 3900

Qy      3901 GTGTTCTTCGGCAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTG 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3901 GTGTTCTTCGGCAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTG 3960

Qy      3961 CTTGACAACATCTATCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGTACAGA 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3961 CTTGACAACATCTATCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGTACAGA 4020

Qy      4021 GTGATCAGAGGTGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGC 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4021 GTGATCAGAGGTGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGC 4080

Qy      4081 AAAGGTCAACCAGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTT 4140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4081 AAAGGTCAACCAGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTT 4140

Qy      4141 GATAGACAGCCAATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACCGCTCATCATA 4200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4141 GATAGACAGCCAATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACCGCTCATCATA 4200

Qy      4201 CATGCTGTAGGACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTAAGCTCGCA 4260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4201 CATGCTGTAGGACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTAAGCTCGCA 4260

Qy      4261 GCTGCCTACATGAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTA 4320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4261 GCTGCCTACATGAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTA 4320

Qy      4321 CCGCTACTGTCAACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCAT 4380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4321 CCGCTACTGTCAACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCAT 4380

Qy      4381 CACCTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAA 4440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4381 CACCTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAA 4440

Qy      4441 CAATGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATTCTGGAT 4500
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4441 CAATGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATTCTGGAT 4500

Qy       4501 GATGACAAGCCAGTAGACATTGACTTGGTCAGGGTCCACCCAAACAGCTCTTTGGCAGGC 4560
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4501 GATGACAAGCCAGTAGACATTGACTTGGTCAGGGTCCACCCAAACAGCTCTTTGGCAGGC 4560

Qy       4561 AGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTGGAAGGTACACGATTC 4620
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4561 AGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTGGAAGGTACACGATTC 4620

Qy       4621 CATCAGACCGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCCAACAAATCTGAGGCT 4680
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4621 CATCAGACCGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCCAACAAATCTGAGGCT 4680

Qy       4681 AATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGC 4740
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4681 AATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGC 4740

Qy       4741 CCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTAC 4800
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4741 CCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTAC 4800

Qy       4801 GCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTA 4860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4801 GCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTA 4860

Qy       4861 TGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGC 4920
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4861 TGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGC 4920

Qy       4921 AAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGCGGAA 4980
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4921 AAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGCGGAA 4980

Qy       4981 ATAATTCTAGAAACGCCACC 5000
                ||||||||||||||||||||
Db       4981 ATAATTCTAGAAACGCCACC 5000


RESULT 2
US-09-991-258-1
; Sequence 1, Application US/09991258
; Patent No. US20020141975A1
; GENERAL INFORMATION:
;  APPLICANT: Olmsted, Robert
;  APPLICANT:  Keith, Paula
;  APPLICANT:  Dryga, Sergey
;  APPLICANT:  Caley, Ian
;  APPLICANT:  Maughan, Maureen
;  APPLICANT:  Johnston, Robert
;  APPLICANT:  Davis, Nancy
;  APPLICANT:  Swanstrom, Ronald
;  TITLE OF INVENTION: ALPHAVIRUS VECTORS AND VIROSOMES WITH MODIFIED HIV GENES FOR USE
;  TITLE OF INVENTION:  VACCINES
;  FILE REFERENCE: 01113.0001U3
;  CURRENT APPLICATION NUMBER: US/09/991,258
;  CURRENT FILING DATE:  2001-11-16
;  PRIOR APPLICATION NUMBER: 09/902,537
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: 60/216,995
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 12523
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence; No. US20020141975A1e =
;   OTHER INFORMATION: synthetic construct
US-09-991-258-1
```

```
  Query Match             42.3%;  Score 2116.4;  DB 10;  Length 12523;
  Best Local Similarity   64.5%;  Pred. No. 0;
  Matches 3218;  Conservative    0;  Mismatches 1746;  Indels   24;  Gaps    3;

Qy         10 ACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTC 69
              | | | | | |||| ||||| | | ||||||||||| | || |  |||||||| |    ||
Db         30 AATTACCTACCCAAAATGGAGAAAGTTCACGTTGACATCGAGGAAGACAGCCCATTCCTC 89

Qy         70 AAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAAT 129
              |   | || |||||||  || || || |||||| | ||||| | ||||||||||| |||
Db         90 AGAGCTTTGCAGCGGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCACTGATAAT 149

Qy        130 GACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTC 189
              |||||||| |||||||||||||||||||| ||||  |||| || || || |  || ||
Db        150 GACCATGCTAATGCCAGAGCGTTTTCGCATCTGGCTTCAAAACTGATCGAAACGGAGGTG 209

Qy        190 GACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAAT 249
              ||||    | |    ||| | ||||||||||||||||||| | |        |||| ||
Db        210 GACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCCGCCCGCAGAATGTATTCTAAG 269

Qy        250 CACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGG 309
              |||   |||||||||||||| || |||| | | || ||||| |||||||| | |  |
Db        270 CACAAGTATCATTGTATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGATTGTATAAG 329

Qy        310 TATGCAGAAAGACTTAAGAAAAGT------GACATTACCGACAAGAACATAGCCTCTAAG 363
              ||||||    |  || |||||||          || || || ||| |  | | |  ||
Db        330 TATGCAACTAAGCTGAAGAAAAACTGTAAGGAAATAACTGATAAGGAATTGGACAAGAAA 389

Qy        364 GCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATG 423
                 |  || ||    | ||||||       || |||  || |||    || ||||  |
Db        390 ATGAAGGAGCTCGCCGCCGTCATGAGCGACCCTGACCTGGAAACTGAGACTATGTGCCTC 449

Qy        424 CACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCA 483
              |||  ||||  |||||| | |||   ||  || || || ||||| ||||| |||||
Db        450 CACGACGACGAGTCGTGTCGCTACGAAGGGCAAGTCGCTGTTTACCAGGATGTATACGCG 509

Qy        484 GTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGG 543
              ||  | | |||||||    |||| |||||  ||   ||| ||||||||       ||||||
Db        510 GTTGACGGACCGACAAGTCTCTATCACCAAGCCAATAAGGGAGTTAGAGTCGCCTACTGG 569

Qy        544 ATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTAC 603
              |||||||||||||| ||||||||||||   || ||| |||| ||  | || ||  | |||
Db        570 ATAGGCTTTGACACCACCCCTTTTATGTTTAAGAACTTGGCTGGAGCATATCCATCATAC 629

Qy        604 AACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCA 663
                 || |||||||| ||||| |  || ||    || |||||||| |||||  | | |||
Db        630 TCTACCAACTGGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCCTATGCAGCTCT 689

Qy        664 GATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCT 723
              ||  ||  |||| |     | | ||     | || || ||||| ||||||     |  ||||
Db        690 GACGTTATGGAGCGGTCACGTAGAGGGATGTCCATTCTTAGAAAGAAGTATTTGAAACCA 749

Qy        724 ACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTA 783
              | || ||  |  |||||||  ||||| || || ||||||   ||  | ||      | |
Db        750 TCCAACAATGTTCTATTCTCTGTTGGCTCGACCATCTACCACGAGAAGAGGGACTTACTG 809

Qy        784 CGTAGCTGGCATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGA 843
              | |||||||| || ||    || || |||||    || |||   || | |||| || |
Db        810 AGGAGCTGGCACCTGCCGTCTGTATTTCACTTACGTGGCAAGCAAAATTACACATGTCGG 869

Qy        844 TGTGGGACCATTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGA 903
              |||| ||| || || || || || |||||||||  | ||||  ||| | ||||| |||||
Db        870 TGTGAGACTATAGTTAGTTGCGACGGGTACGTCGTTAAAAGAATAGCTATCAGTCCAGGC 929

Qy        904 CTATACGGTAAAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGC 963
              || || || ||    |     ||  ||  | || ||||| |||||||| |||||| | |||
Db        930 CTGTATGGGAAGCCTTCAGGCTATGCTGCTACGATGCACCGCGAGGGATTCTTGTGCTGC 989

Qy        964 AAAGTCACAGATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCA 1023
              ||||| ||||| ||  |   ||| |||||||| |||||| | ||||| |||||||| |||
Db        990 AAAGTGACAGACACATTAAACGGGGAGAGGGTCTCTTTTCCCGTGTGCACGTATGTGCCA 1049

Qy       1024 GCCACACTTTGCGATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCA 1083
```


Db 1050 GCTACATTGTGTGACCAAATGACTGGCATACTGGCAACAGATGTCAGTGCGGACGACGCG 1109
Qy 1084 CAAAAACTATTGGTTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAAT 1143
Db 1110 CAAAAACTGCTGGTTGGGCTCAACCAGCGTATAGTCGTCAACGGTCGCACCCAGAGAAAC 1169
Qy 1144 ACTAACACAATGCAGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCG 1203
Db 1170 ACCAATACCATGAAAAATTACCTTTTGCCCGTAGTGGCCCAGGCATTTGCTAGGTGGGCA 1229
Qy 1204 CGTGAACATCGTGCCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTT 1263
Db 1230 AAGGAATATAAGGAAGATCAAGAAGATGAAAGGCCACTAGGACTACGAGATAGACAGTTA 1289
Qy 1264 ACTATGGGCTGCTGCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCT 1323
Db 1290 GTCATGGGGTGTTGTTGGGCTTTTAGAAGGCACAAGATAACATCTATTTATAAGCGCCCG 1349
Qy 1324 GGTACGCAAACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTT 1383
Db 1350 GATACCCAAACCATCATCAAAGTGAACAGCGATTTCCACTCATTCGTGCTGCCCAGGATA 1409
Qy 1384 ACCAGCCACGGGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACT 1443
Db 1410 GGCAGTAACACATTGGAGATCGGGCTGAGAACAAGAATCAGGAAAATGTTAGAGGAGCAC 1469
Qy 1444 GTCAAACCCGCACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAA 1503
Db 1470 AAGGAGCCGTCACCTCTCATTACCGCCGAGGACGTACAAGAAGCTAAGTGCGCAGCCGAT 1529
Qy 1504 GAAGCTGAAGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCT 1563
Db 1530 GAGGCTAAGGAGGTGCGTGAAGCCGAGGAGTTGCGCGCAGCTCTACCACCTTTGGCAGCT 1589
Qy 1564 GAAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGAGCA 1623
Db 1590 GATGTTGAGGAGCCCACTCTGGAAGCCGATGTCGACTTGATGTTACAAGAGGCTGGGGCC 1649
Qy 1624 GGTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAG 1683
Db 1650 GGCTCAGTGGAGACACCTCGTGGCTTGATAAAGGTTACCAGCTACGCTGGCGAGGACAAG 1709
Qy 1684 ATTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGT 1743
Db 1710 ATCGGCTCTTACGCTGTGCTTTCTCCGCAGGCTGTACTCAAGAGTGAAAAATTATCTTGC 1769
Qy 1744 ATCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATAC 1803
Db 1770 ATCCACCCTCTCGCTGAACAAGTCATAGTGATAACACACTCTGGCCGAAAAGGGCGTTAT 1829
Qy 1804 AAAGTCGAGCCATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAA 1863
Db 1830 GCCGTGGAACCATACCATGGTAAAGTAGTGGTGCCAGAGGGACATGCAATACCCGTCCAG 1889
Qy 1864 GACTTCCAGGCATTGAGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAAC 1923
Db 1890 GACTTTCAAGCTCTGAGTGAAAGTGCCACCATTGTGTACAACGAACGTGAGTTCGTAAAC 1949
Qy 1924 AGATACCTGCACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTAT 1983
Db 1950 AGGTACCTGCACCATATTGCCACACATGGAGGAGCGCTGAACACTGATGAAGAATATTAC 2009
Qy 1984 AAGACTGTAAAGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAG 2043
Db 2010 AAAACTGTCAAGCCCAGCGAGCACGACGGCGAATACCTGTACGACATCGACAGGAAACAG 2069
Qy 2044 TGTGTTAAGCGAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCA 2103
Db 2070 TGCGTCAAGAAAGAACTAGTCACTGGGCTAGGGCTCACAGGCGAGCTGGTGGATCCTCCC 2129
Qy 2104 TTTCACGAGTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCA 2163
Db 2130 TTCCATGAATTCGCCTACGAGAGTCTGAGAACACGACCAGCCGCTCCTTACCAAGTACCA 2189
Qy 2164 ACCATCGGAGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTG 2223
Db 2190 ACCATAGGGGTGTATGGCGTGCCAGGATCAGGCAAGTCTGGCATCATTAAAAGCGCAGTC 2249
Qy 2224 ACTAAGAAAGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGAT 2283
Db 2250 ACCAAAAAAGATCTAGTGGTGAGCGCCAAGAAAGAAAACTGTGCAGAAATTATAAGGGAC 2309
Qy 2284 GTAAGGAGGATGAGACGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAAT 2343
Db 2310 GTCAAGAAAATGAAAGGGCTGGACGTCAATGCCAGAACTGTGGACTCAGTGCTCTTGAAT 2369
Qy 2344 GGGGTTAAGCACCCCGTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGG 2403
Db 2370 GGATGCAAACACCCCGTAGAGACCCTGTATATTGACGAAGCTTTTGCTTGTCATGCAGGT 2429
Qy 2404 ACGCTGCTGGCACTGATTGCCATCGTCAAACCTAAGAAAGTGGTATTGTGCGGGGACCCA 2463
Db 2430 ACTCTCAGAGCGCTCATAGCCATTATAAGACCTAAAAAGGCAGTGCTCTGCGGGGATCCC 2489
Qy 2464 AAACAATGCGGCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGC 2523
Db 2490 AAACAGTGCGGTTTTTTTAACATGATGTGCCTGAAAGTGCATTTTAACCACGAGATTTGC 2549
Qy 2524 ACTGAAGTGTACCATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATCGTC 2583
Db 2550 ACACAAGTCTTCCACAAAAGCATCTCTCGCCGTTGCACTAAATCTGTGACTTCGGTCGTC 2609
Qy 2584 TCCACGCTCTTCTACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATCATC 2643
Db 2610 TCAACCTTGTTTTACGACAAAAAAATGAGAACGACGAATCCGAAAGAGACTAAGATTGTG 2669
Qy 2644 ATAGATACCACAGGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGA 2703
Db 2670 ATTGACACTACCGGCAGTACCAAACCTAAGCAGGACGATCTCATTCTCACTTGTTTCAGA 2729
Qy 2704 GGATGGGTGAAACAGCTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCTGCA 2763
Db 2730 GGGTGGGTGAAGCAGTTGCAAATAGATTACAAAGGCAACGAAATAATGACGGCAGCTGCC 2789
Qy 2764 TCGCAAGGACTTACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCA 2823
Db 2790 TCTCAAGGGCTGACCCGTAAAGGTGTGTATGCCGTTCGGTACAAGGTGAATGAAAATCCT 2849
Qy 2824 CTCTACTCGCAGACTTCTGAGCACGTGAACGTGTTACTTACACGCACAGAAAAACGCATT 2883
Db 2850 CTGTACGCACCCACCTCAGAACATGTGAACGTCCTACTGACCCGCACGGAGGACCGCATC 2909
Qy 2884 GTCTGGAAGACGCTAGCTGGTGATCCCTGGATAAAGACACTTACAGCTAAATATCCCGGG 2943
Db 2910 GTGTGGAAAACACTAGCCGGCGACCCATGGATAAAAACACTGACTGCCAAGTACCCTGGG 2969
Qy 2944 GATTTCACGGCTTCATTGGACGACTGGCAGCGCGAACACGACGCCATTATGGCACGCGTT 3003
Db 2970 AATTTCACTGCCACGATAGAGGAGTGGCAAGCAGAGCATGATGCCATCATGAGGCACATC 3029
Qy 3004 CTTGATAAGCCGCAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAG 3063
Db 3030 TTGGAGAGACCGGACCCTACCGACGTCTTCCAGAATAAGGCAAACGTGTGTTGGGCCAAG 3089
Qy 3064 GCTTTAGAGCCAGTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAGACG 3123
Db 3090 GCTTTAGTGCCGGTGCTGAAGACCGCTGGCATAGACATGACCACTGAACAATGGAACACT 3149
Qy 3124 TTGCACCCATTCAAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTTTGC 3183
Db 3150 GTGGATTATTTTGAAACGGACAAAGCTCACTCAGCAGAGATAGTATTGAACCAACTATGC 3209
Qy 3184 ACCAGGTTCTTTGGAGTAGACCTGGACAGTGGGTTATTTTCCGCTCCTACCGTCGCACTT 3243
Db 3210 GTGAGGTTCTTTGGACTCGATCTGGACTCCGGTCTATTTTCTGCACCCACTGTTCCGTTA 3269

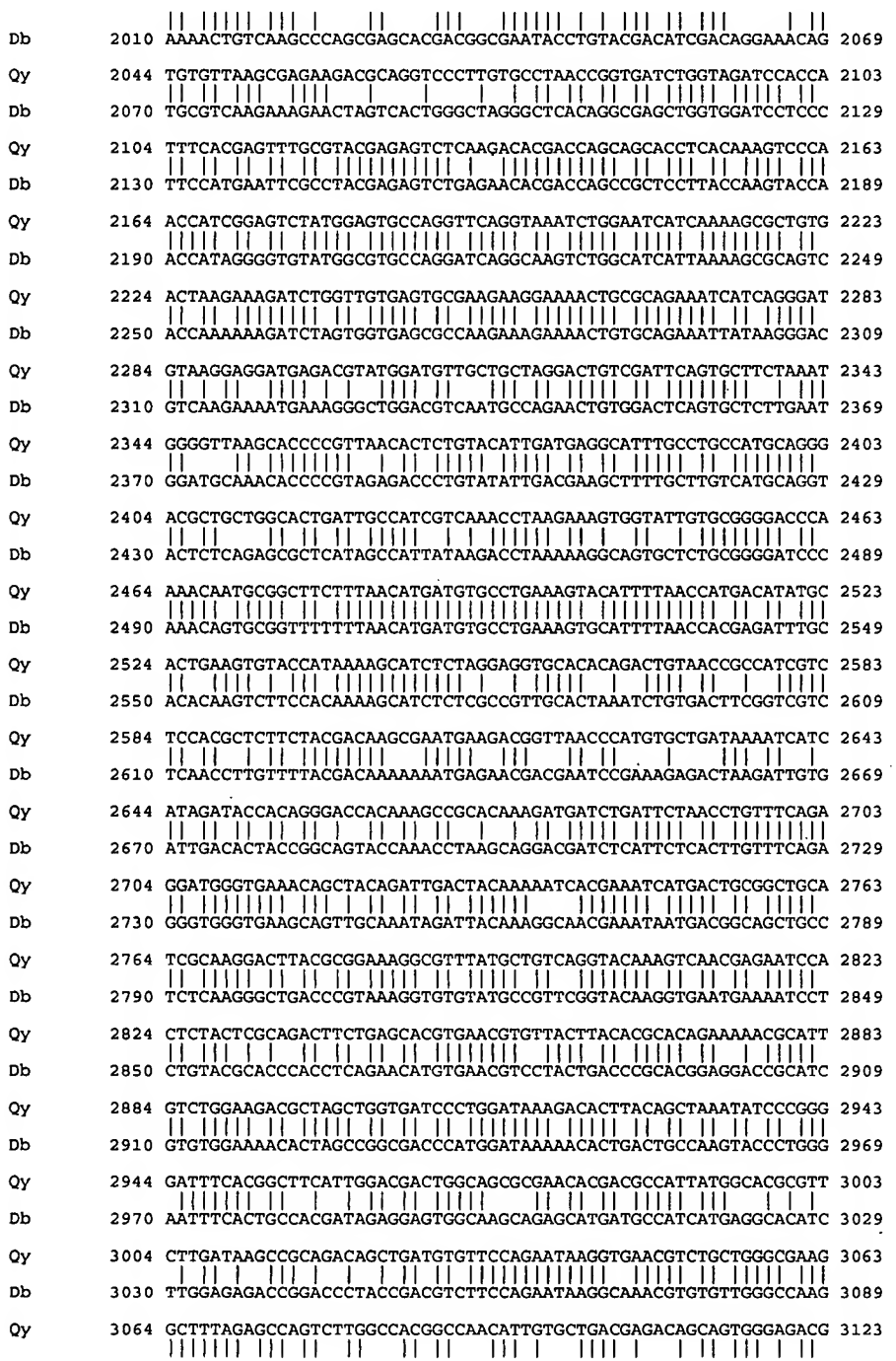


RESULT 3
US-09-991-258-14
; Sequence 14, Application US/09991258
; Patent No. US20020141975A1
; GENERAL INFORMATION:
; APPLICANT: Olmsted, Robert
; APPLICANT: Keith, Paula
; APPLICANT: Dryga, Sergey
; APPLICANT: Caley, Ian
; APPLICANT: Maughan, Maureen
; APPLICANT: Johnston, Robert
; APPLICANT: Davis, Nancy
; APPLICANT: Swanstrom, Ronald
; TITLE OF INVENTION: ALPHAVIRUS VECTORS AND VIROSOMES WITH MODIFIED HIV GENES FOR USE I
; TITLE OF INVENTION: VACCINES
; FILE REFERENCE: 01113.0001U3
; CURRENT APPLICATION NUMBER: US/09/991,258
; CURRENT FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: 09/902,537
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: 60/216,995
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 4.0

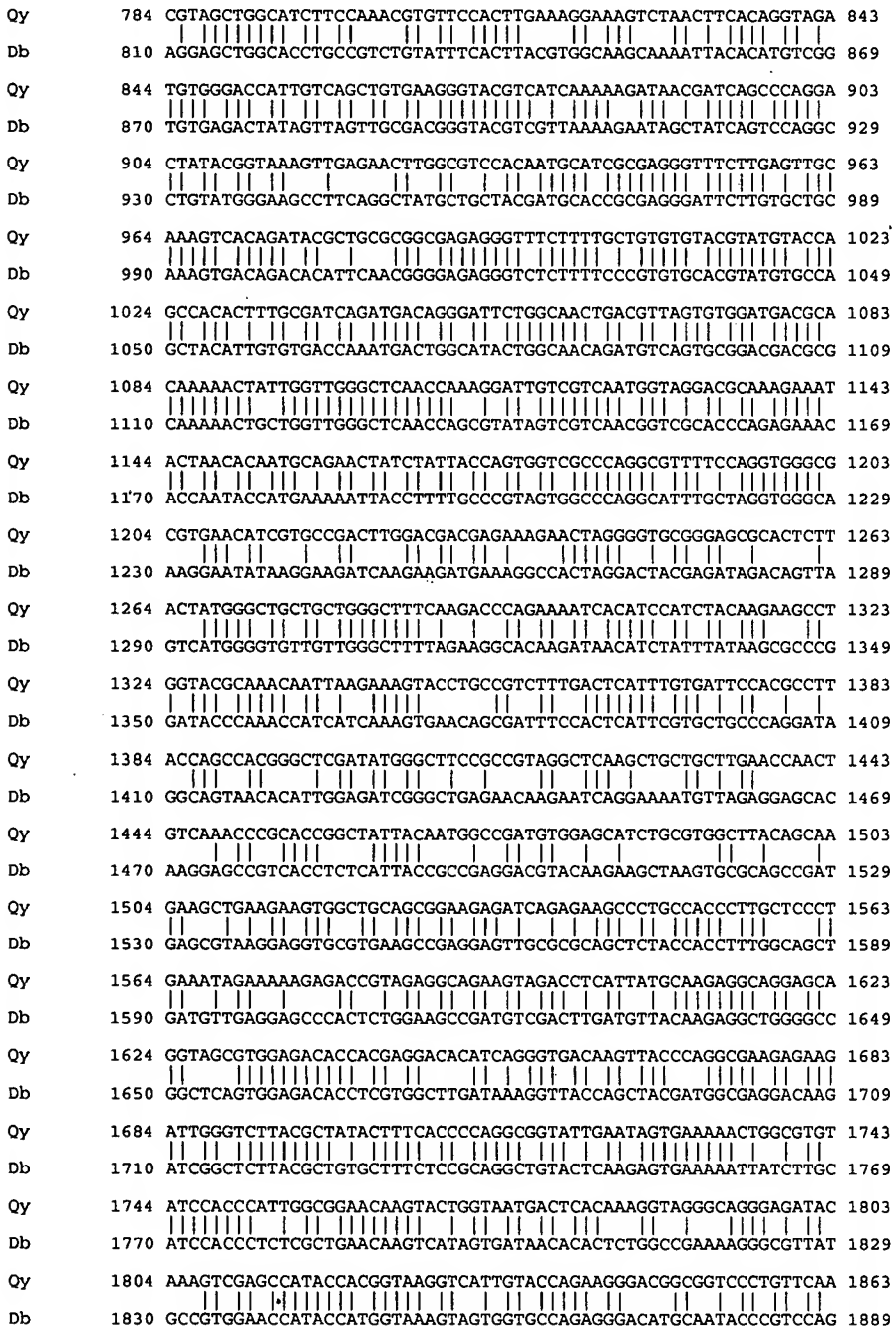
; SEQ ID NO 14
; LENGTH: 12379
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. US20020141975A1e =
; OTHER INFORMATION: synthetic construct
US-09-991-258-14

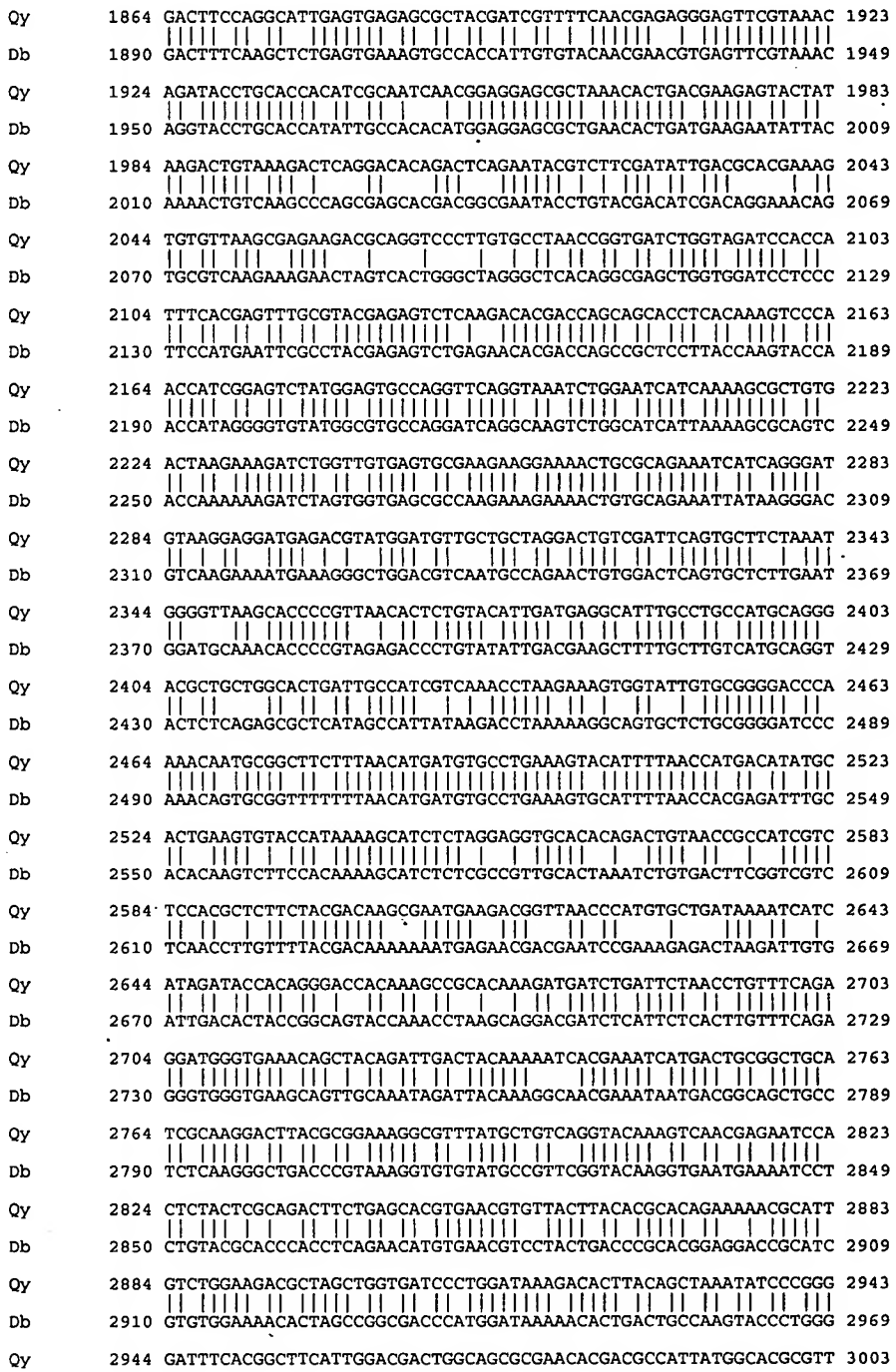
Query Match 42.3%; Score 2113.2; DB 10; Length 12379;
Best Local Similarity 64.5%; Pred. No. 0;
Matches 3216; Conservative 0; Mismatches 1748; Indels 24; Gaps 3;

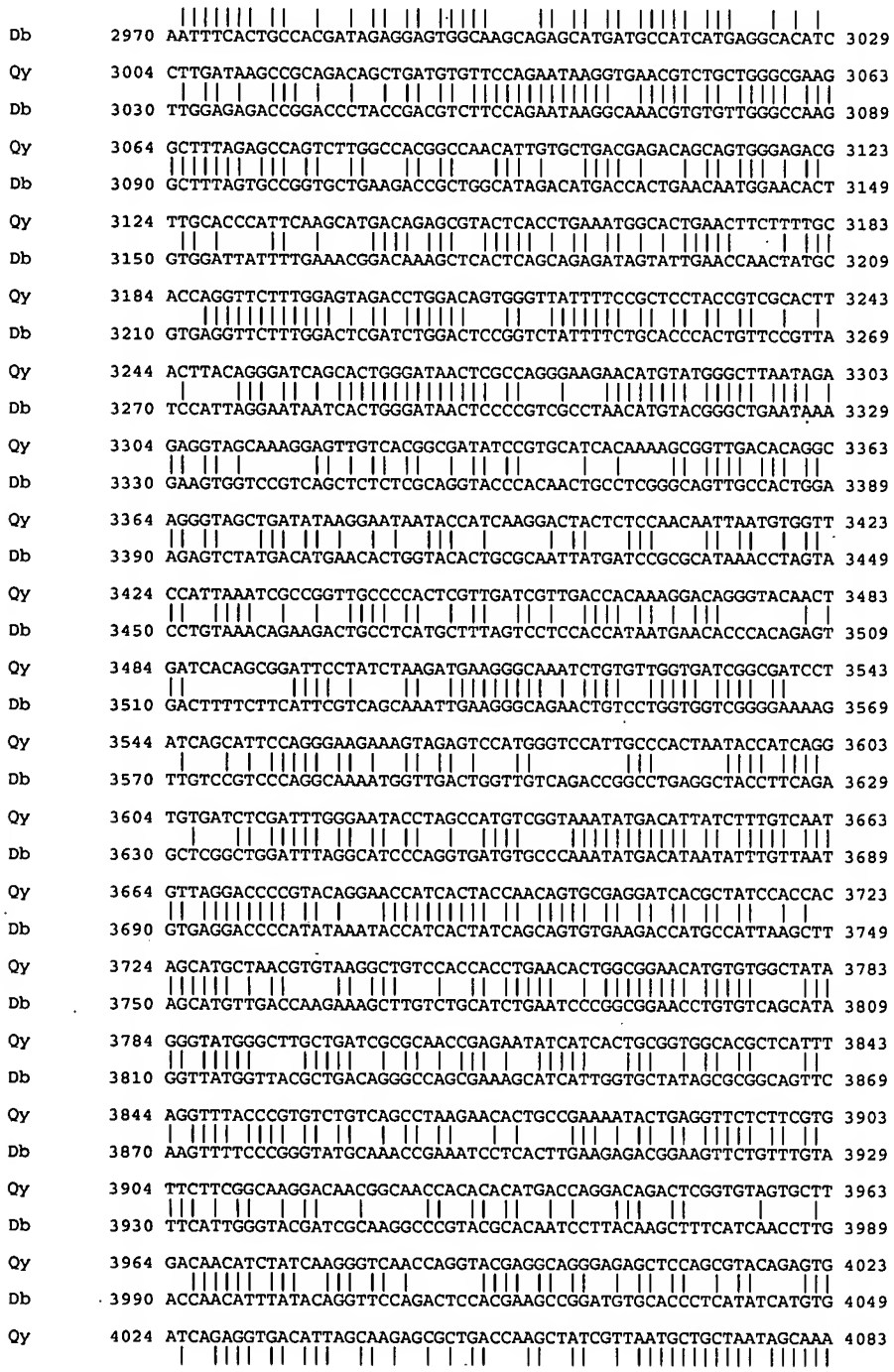
```
Qy      10 ACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTC 69
           | | | | | |||| ||||| | | |||||||||||| | || | |||||||| |    ||
Db      30 AATTACCTACCCAAAATGGAGAAAGTTCACGTTGACATCGAGGAAGACAGCCCATTCCTC 89

Qy      70 AAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAAT 129
           |    | || |||||||  || || || |||||| | ||||| | ||||||||||| |||
Db      90 AGAGCTTTGCAGCGGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCACTGATAAT 149

Qy     130 GACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTC 189
           |||||||| |||||||||||||||||||| ||||  |||| || || || |  || ||
Db     150 GACCATGCTAATGCCAGAGCGTTTTCGCATCTGGCTTCAAAACTGATCGAAACGGAGGTG 209

Qy     190 GACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAAT 249
           ||||    | |    ||| | |||||||||||||||||||| | |        |||| ||
Db     210 GACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCCGCCCGCAGAATGTATTCTAAG 269

Qy     250 CACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGG 309
           |||   |||||||||||||| || |||| | | || ||||| |||||||| |  |   |
Db     270 CACAAGTATCATTGTATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGATTGTATAAG 329

Qy     310 TATGCAGAAAGACTTAAGAAAAGT------GACATTACCGACAAGAACATAGCCTCTAAG 363
           ||||||   |  || |||||||         || || || || ||| |  | | |   ||
Db     330 TATGCAACTAAGCTGAAGAAAAACTGTAAGGAAATAACTGATAAGGAATTGGACAAGAAA 389

Qy     364 GCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATG 423
           |   || ||    |  ||||||       || |||   || |||    || ||||  |
Db     390 ATGAAGGAGCTCGCCGCCGTCATGAGCGACCCTGACCTGGAAACTGAGACTATGTGCCTC 449

Qy     424 CACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCA 483
           |||   ||||   ||||| | |||   ||    || || || ||||| ||||| |||||
Db     450 CACGACGACGAGTCGTGTCGCTACGAAGGGCAAGTCGCTGTTTACCAGGATGTATACGCG 509

Qy     484 GTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGG 543
           ||  | | |||||||     |||| ||||| ||   ||| ||||||||        ||||||
Db     510 GTTGACGGACCGACAAGTCTCTATCACCAAGCCAATAAGGGAGTTAGAGTCGCCTACTGG 569

Qy     544 ATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTAC 603
           |||||||||||||| ||||||||||||   || ||| |||| ||  | || ||  | |||
Db     570 ATAGGCTTTGACACCACCCCTTTTATGTTTAAGAACTTGGCTGGAGCATATCCATCATAC 629

Qy     604 AACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCA 663
             || |||||||| ||||| |  || ||    || |||||||| |||||  | | |||
Db     630 TCTACCAACTGGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCCTATGCAGCTCT 689

Qy     664 GATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCT 723
           ||  ||  |||| |    | | ||      | || || ||||| |||||     |  ||||
Db     690 GACGTTATGGAGCGGTCACGTAGAGGGATGTCCATTCTTAGAAAGAAGTATTTGAAACCA 749

Qy     724 ACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTA 783
           | || ||  |  ||||||| ||||| || || ||||||   ||  | ||    |  |
Db     750 TCCAACAATGTTCTATTCTCTGTTGGCTCGACCATCTACCACGAGAAGAGGGACTTACTG 809

Qy     784 CGTAGCTGGCATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGA 843
           | ||||||| || ||    || || |||||    || |||   || | |||| || |
Db     810 AGGAGCTGGCACCTGCCGTCTGTATTTCACTTACGTGGCAAGCAAAATTACACATGTCGG 869

Qy     844 TGTGGGACCATTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGA 903
           |||| ||| || || || || || |||||||||| | |||| ||| | ||||| |||||
Db     870 TGTGAGACTATAGTTAGTTGCGACGGGTACGTCGTTAAAAGAATAGCTATCAGTCCAGGC 929

Qy     904 CTATACGGTAAAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGC 963
           || || || ||    |      ||  ||  | || ||||| |||||||| ||||||  | |||
Db     930 CTGTATGGGAAGCCTTCAGGCTATGCTGCTACGATGCACCGCGAGGGATTCTTGTGCTGC 989

Qy     964 AAAGTCACAGATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCA 1023
           ||||| ||||| ||   |   ||| |||||||| |||||| | |||||  |||||||| |||
Db     990 AAAGTGACAGACACATTCAACGGGGAGAGGGTCTCTTTTCCCGTGTGCACGTATGTGCCA 1049

Qy    1024 GCCACACTTTGCGATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCA 1083
           || ||| | || || || ||||| || || |||||||| || || |||| ||| |||||
Db    1050 GCTACATTGTGTGACCAAATGACTGGCATACTGGCAACAGATGTCAGTGCGGACGACGCG 1109

Qy    1084 CAAAAACTATTGGTTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAAT 1143
           ||||||||  |||||||||||||||||  | || |||||||| ||| | || || |||||
Db    1110 CAAAAACTGCTGGTTGGGCTCAACCAGCGTATAGTCGTCAACGGTCGCACCCAGAGAAAC 1169

Qy    1144 ACTAACACAATGCAGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCG 1203
           || || || ||| | || || || || || || || || |||||||| ||| | ||||||||
Db    1170 ACCAATACCATGAAAAATTACCTTTTGCCCGTAGTGGCCCAGGCATTTGCTAGGTGGGCA 1229

Qy    1204 CGTGAACATCGTGCCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTT 1263
           |||  ||   |  ||    || || || |    ||||||  | || ||  |      |
Db    1230 AAGGAATATAAGGAAGATCAAGAAGATGAAAGGCCACTAGGACTACGAGATAGACAGTTA 1289

Qy    1264 ACTATGGGCTGCTGCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCT 1323
           |||||| || || |||||||| |  | || || || ||||| || || |||    ||
Db    1290 GTCATGGGGTGTTGTTGGGCTTTTAGAAGGCACAAGATAACATCTATTTATAAGCGCCCG 1349

Qy    1324 GGTACGCAAACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTT 1383
           | ||| ||||| || |  |||||        ||  ||  ||||||| ||| | ||  |  |
Db    1350 GATACCCAAACCATCATCAAAGTGAACAGCGATTTCCACTCATTCGTGCTGCCCAGGATA 1409

Qy    1384 ACCAGCCACGGGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACT 1443
           |||  ||    | || || ||  |  |    ||  |||  |    || | ||
Db    1410 GGCAGTAACACATTGGAGATCGGGCTGAGAACAAGAATCAGGAAAATGTTAGAGGAGCAC 1469

Qy    1444 GTCAAACCCGCACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAA 1503
           .  | ||  ||||   |||||     |  || ||  |  |          ||  |    |
Db    1470 AAGGAGCCGTCACCTCTCATTACCGCCGAGGACGTACAAGAAGCTAAGTGCGCAGCCGAT 1529

Qy    1504 GAAGCTGAAGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCT 1563
           ||   | | || |||  || ||| || |||  |  | ||| || ||||| |||    ||
Db    1530 GAGCGTAAGGAGGTGCGTGAAGCCGAGGAGTTGCGCGCAGCTCTACCACCTTTGGCAGCT 1589

Qy    1564 GAAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGAGCA 1623
           ||  | ||  |    ||  | || || || || ||| | ||  | |||||||| || ||
Db    1590 GATGTTGAGGAGCCCACTCTGGAAGCCGATGTCGACTTGATGTTACAAGAGGCTGGGGCC 1649

Qy    1624 GGTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAG 1683
           ||     |||||||||| || ||    || | ||| || || |||   ||||| || |||
Db    1650 GGCTCAGTGGAGACACCTCGTGGCTTGATAAAGGTTACCAGCTACGATGGCGAGGACAAG 1709

Qy    1684 ATTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGT 1743
           || || |||||||| | ||||| || ||||| ||| | || |||||||| |  | ||
Db    1710 ATCGGCTCTTACGCTGTGCTTTCTCCGCAGGCTGTACTCAAGAGTGAAAAATTATCTTGC 1769

Qy    1744 ATCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATAC 1803
           ||||||||  | || |||||||| | || || || |||   ||  |  |||| | ||
Db    1770 ATCCACCCTCTCGCTGAACAAGTCATAGTGATAACACACTCTGGCCGAAAAGGGCGTTAT 1829

Qy    1804 AAAGTCGAGCCATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAA 1863
           ||  || |||||||| ||||| || |  || |||||  ||    || | || || ||
Db    1830 GCCGTGGAACCATACCATGGTAAAGTAGTGGTGCCAGAGGGACATGCAATACCCGTCCAG 1889

Qy    1864 GACTTCCAGGCATTGAGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAAC 1923
           |||||  || ||  ||||||| || || || || || | ||||||  | ||||||||||||
Db    1890 GACTTTCAAGCTCTGAGTGAAAGTGCCACCATTGTGTACAACGAACGTGAGTTCGTAAAC 1949

Qy    1924 AGATACCTGCACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTAT 1983
           || |||||||||| || || |   | |||||||||| |||||||| ||||| || ||
Db    1950 AGGTACCTGCACCATATTGCCACACATGGAGGAGCGCTGAACACTGATGAAGAATATTAC 2009

Qy    1984 AAGACTGTAAAGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAG 2043
```


Db 2010 AAAACTGTCAAGCCCAGCGAGCACGACGGCGAATACCTGTACGACATCGACAGGAAACAG 2069
Qy 2044 TGTGTTAAGCGAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCA 2103
Db 2070 TGCGTCAAGAAAGAACTAGTCACTGGGCTAGGGCTCACAGGCGAGCTGGTGGATCCTCCC 2129
Qy 2104 TTTCACGAGTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCA 2163
Db 2130 TTCCATGAATTCGCCTACGAGAGTCTGAGAACACGACCAGCCGCTCCTTACCAAGTACCA 2189
Qy 2164 ACCATCGGAGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTG 2223
Db 2190 ACCATAGGGGTGTATGGCGTGCCAGGATCAGGCAAGTCTGGCATCATTAAAAGCGCAGTC 2249
Qy 2224 ACTAAGAAAGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGAT 2283
Db 2250 ACCAAAAAAGATCTAGTGGTGAGCGCCAAGAAAGAAAACTGTGCAGAAATTATAAGGGAC 2309
Qy 2284 GTAAGGAGGATGAGACGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAAT 2343
Db 2310 GTCAAGAAAATGAAAGGGCTGGACGTCAATGCCAGAACTGTGGACTCAGTGCTCTTGAAT 2369
Qy 2344 GGGGTTAAGCACCCCGTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGG 2403
Db 2370 GGATGCAAACACCCCGTAGAGACCCTGTATATTGACGAAGCTTTGCTTGTCATGCAGGT 2429
Qy 2404 ACGCTGCTGGCACTGATTGCCATCGTCAAACCTAAGAAAGTGGTATTGTGCGGGGACCCA 2463
Db 2430 ACTCTCAGAGCGCTCATAGCCATTATAAGACCTAAAAAGGCAGTGCTCTGCGGGGATCCC 2489
Qy 2464 AAACAATGCGGCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGC 2523
Db 2490 AAACAGTGCGGTTTTTTTAACATGATGTGCCTGAAAGTGCATTTTAACCACGAGATTTGC 2549
Qy 2524 ACTGAAGTGTACCATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATCGTC 2583
Db 2550 ACACAAGTCTTCCACAAAAGCATCTCTCGCCGTTGCACTAAATCTGTGACTTCGGTCGTC 2609
Qy 2584 TCCACGCTCTTCTACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATCATC 2643
Db 2610 TCAACCTTGTTTTACGACAAAAAAATGAGAACGACGAATCCGAAAGAGACTAAGATTGTG 2669
Qy 2644 ATAGATACCACAGGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGA 2703
Db 2670 ATTGACACTACCGGCAGTACCAAACCTAAGCAGGACGATCTCATTCTCACTTGTTTCAGA 2729
Qy 2704 GGATGGGTGAAACAGCTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCTGCA 2763
Db 2730 GGGTGGGTGAAGCAGTTGCAAATAGATTACAAAGGCAACGAAATAATGACGGCAGCTGCC 2789
Qy 2764 TCGCAAGGACTTACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCA 2823
Db 2790 TCTCAAGGGCTGACCCGTAAAGGTGTGTATGCCGTTCGGTACAAGGTGAATGAAAATCCT 2849
Qy 2824 CTCTACTCGCAGACTTCTGAGCACGTGAACGTGTTACTTACACGCACAGAAAAACGCATT 2883
Db 2850 CTGTACGCACCCACCTCAGAACATGTGAACGTCCTACTGACCCGCACGGAGGACCGCATC 2909
Qy 2884 GTCTGGAAGACGCTAGCTGGTGATCCCTGGATAAAGACACTTACAGCTAAATATCCCGGG 2943
Db 2910 GTGTGGAAAACACTAGCCGGCGACCCATGGATAAAAACACTGACTGCCAAGTACCCTGGG 2969
Qy 2944 GATTTCACGGCTTCATTGGACGACTGGCAGCGCGAACACGACGCCATTATGGCACGCGTT 3003
Db 2970 AATTTCACTGCCACGATAGAGGAGTGGCAAGCAGAGCATGATGCCATCATGAGGCACATC 3029
Qy 3004 CTTGATAAGCCGCAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAG 3063
Db 3030 TTGGAGAGACCGGACCCTACCGACGTCTTCCAGAATAAGGCAAACGTGTGTTGGGCCAAG 3089
Qy 3064 GCTTTAGAGCCAGTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAGACG 3123
Db 3090 GCTTTAGTGCCGGTGCTGAAGACCGCTGGCATAGACATGACCACTGAACAATGGAACACT 3149
Qy 3124 TTGCACCCATTCAAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTTTGC 3183
Db 3150 GTGGATTATTTTGAAACGGACAAAGCTCACTCAGCAGAGATAGTATTGAACCAACTATGC 3209
Qy 3184 ACCAGGTTCTTTGGAGTAGACCTGGACAGTGGGTTATTTTCCGCTCCTACCGTCGCACTT 3243
Db 3210 GTGAGGTTCTTTGGACTCGATCTGGACTCCGGTCTATTTTCTGCACCCACTGTTCCGTTA 3269
Qy 3244 ACTTACAGGGATCAGCACTGGGATAACTCGCCAGGGAAGAACATGTATGGGCTTAATAGA 3303
Db 3270 TCCATTAGGAATAATCACTGGGATAACTCCCCGTCGCCTAACATGTACGGGCTGAATAAA 3329
Qy 3304 GAGGTAGCAAAGGAGTTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACAGGC 3363
Db 3330 GAAGTGGTCCGTCAGCTCTCTCGCAGGTACCCACAACTGCCTCGGGCAGTTGCCACTGGA 3389
Qy 3364 AGGGTAGCTGATATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTGGTT 3423
Db 3390 AGAGTCTATGACATGAACACTGGTACACTGCGCAATTATGATCCGCGCATAAACCTAGTA 3449
Qy 3424 CCATTAAATCGCCGGTTGCCCCACTCGTTGATCGTTGACCACAAAGGACAGGGTACAACT 3483
Db 3450 CCTGTAAACAGAAGACTGCCTCATGCTTTAGTCCTCCACCATAATGAACACCCACAGAGT 3509
Qy 3484 GATCACAGCGGATTCCTATCTAAGATGAAGGGCAAATCTGTGTTGGTGATCGGCGATCCT 3543
Db 3510 GACTTTTCTTCATTCGTCAGCAAATTGAAGGGCAGAACTGTCCTGGTGGTCGGGGAAAAG 3569
Qy 3544 ATCAGCATTCCAGGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATACCATCAGG 3603
Db 3570 TTGTCCGTCCCAGGCAAAATGGTTGACTGGTTGTCAGACCGGCCTGAGGCTACCTTCAGA 3629
Qy 3604 TGTGATCTCGATTTGGGAATACCTAGCCATGTCGGTAAATATGACATTATCTTTGTCAAT 3663
Db 3630 GCTCGGCTGGATTTAGGCATCCCAGGTGATGTGCCCAAATATGACATAATATTTGTTAAT 3689
Qy 3664 GTTAGGACCCCGTACAGGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCACCAC 3723
Db 3690 GTGAGGACCCCATATAAATACCATCACTATCAGCAGTGTGAAGACCATGCCATTAAGCTT 3749
Qy 3724 AGCATGCTAACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCTATA 3783
Db 3750 AGCATGTTGACCAAGAAAGCTTGTCTGCATCTGAATCCCGGCGGAACCTGTGTCAGCATA 3809
Qy 3784 GGGTATGGGCTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCACGCTCATTT 3843
Db 3810 GGTTATGGTTACGCTGACAGGGCCAGCGAAAGCATCATTGGTGCTATAGCGCGGCAGTTC 3869
Qy 3844 AGGTTTACCCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTCGTG 3903
Db 3870 AAGTTTTCCCGGGTATGCAAACCGAAATCCTCACTTGAAGAGACGGAAGTTCTGTTTGTA 3929
Qy 3904 TTCTTCGGCAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTGCTT 3963
Db 3930 TTCATTGGGTACGATCGCAAGGCCCGTACGCACAATCCTTACAAGCTTTCATCAACCTTG 3989
Qy 3964 GACAACATCTATCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGTACAGAGTG 4023
Db 3990 ACCAACATTTATACAGGTTCCAGACTCCACGAAGCCGGATGTGCACCCTCATATCATGTG 4049
Qy 4024 ATCAGAGGTGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGCAAA 4083
Db 4050 GTGCGAGGGGATATTGCCACGGCCACCGAAGGAGTGATTATAAATGCTGCTAACAGCAAA 4109
Qy 4084 GGTCAACCAGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTTGAT 4143
Db 4110 GGACAACCTGGCGGAGGGGTGTGCGGAGCGCTGTATAAGAAATTCCCGGAAAGCTTCGAT 4169
Qy 4144 AGACAGCCAATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACC---GCTCATCATA 4200
Db 4170 TTACAGCCGATCGAAGTAGGAAAAGCGCGACTGGTCAAAGGTGCAGCTAAACATATCATT 4229

```
Qy       4201 CATGCTGTAGGACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTAAGCTCGCA 4260
             ||||| |||||||| || ||    ||  |  ||||    || || |||    || | |||
Db       4230 CATGCCGTAGGACCAAACTTCAACAAAGTTTCGGAGGTTGAAGGTGACAAACAGTTGGCA 4289

Qy       4261 GCTGCCTACATGAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTA 4320
             |  || || .  |  ||| ||    || |||||||  | |       |    | || |  |
Db       4290 GAGGCTTATGAGTCCATCGCTAAGATTGTCAACGATAACAATTACAAGTCAGTAGCGATT 4349

Qy       4321 CCGCTACTGTCAACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCAT 4380
             ||  ||  |||| |||||||||| ||| ||   ||||||||||  |  |  |||||||||  |
Db       4350 CCACTGTTGTCCACCGGCATCTTTTCCGGGAACAAAGATCGACTAACCCAATCATTGAAC 4409

Qy       4381 CACCTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAA 4440
             ||  || | || ||||| ||||| || ||||| ||||  |||||||  |||   ||| ||
Db       4410 CATTTGCTGACAGCTTTAGACACCACTGATGCAGATGTAGCCATATACTGCAGGGACAAG 4469

Qy       4441 CAATGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATTCTGGA- 4499
              |||||||  |  |  | |  || ||  |     | | ||||   || ||      |
Db       4470 AAATGGGAAATGACTCTCAAGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGGAGATATGC 4529

Qy       4500 --------------TGATGACAAGCCAGTAGACATTGACTTGGTCAGGGTCCACCCAAAC 4545
                           |   |  |    |   ||   ||  |||| ||||| || || ||
Db       4530 ATATCCGACGACTCTTCAGTGACAGAACCTGATGCAGAGCTGGTGAGGGTGCATCCGAAG 4589

Qy       4546 AGCTCTTTGGCAGGCAGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTG 4605
             || |||||||| || ||    || |||  |   |  || |||||     |  |||||  ||
Db       4590 AGTTCTTTGGCTGGAAGGAAGGGCTACAGCACAAGCGATGGCAAAACTTTCTCATATTTG 4649

Qy       4606 GAAGGTACACGATTCCATCAGACCGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCC 4665
             |||||  ||    || || |||  | |||||||  || || |||||  |||| |||||||||
Db       4650 GAAGGGACCAAGTTTCACCAGGCGGCCAAGGATATAGCAGAAATTAATGCCATGTGGCCC 4709

Qy       4666 AACAAATCTGAGGCTAATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGC 4725
                | | ||||| |||||||||  | |||  ||||  ||||| || || || |||  |||
Db       4710 GTTGCAACGGAGGCCAATGAGCAGGTATGCATGTATATCCTCGGAGAAAGCATGAGCAGT 4769

Qy       4726 ATCCGCTCCAAATGCCCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCA 4785
             ||  | || |||||||| || || |||||  || || ||  | ||||||  |||  || ||
Db       4770 ATTAGGTCGAAATGCCCCGTCGAAGAGTCGGAAGCCTCCTCACCACCTAGCACGCTGCCT 4829

Qy       4786 TGCCTGTGTAATTACGCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAA 4845
             |||  |||| |   | || |||||  | ||  | |||  |  |  |     |  |       |
Db       4830 TGCTTGTGCATCCATGCCATGACTCCAGAAAGAGTACAGCGCCTAAAAGCCTCACGTCCA 4889

Qy       4846 GAACAGTTCGCCGTATGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAG 4905
             |||||  |  | || |||||||| ||  |  |||||||||| || ||||| || ||||||
Db       4890 GAACAAATTACTGTGTGCTCATCCTTTCCATTGCCGAAGTATAGAATCACTGGTGTGCAG 4949

Qy       4906 AAGCTACAGTGCAGCAAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCC 4965
             |||  | || |||  |  |  ||  |  ||| |||      || ||  ||  | | || ||
Db       4950 AAGATCCAATGCTCCCAGCCTATATTGTTCTCACCGAAAGTGCCTGCGTATATTCATCCA 5009

Qy       4966 AGGAAGTA 4973
              ||||||||
Db       5010 AGGAAGTA 5017


RESULT 4
US-09-991-258-17
; Sequence 17, Application US/09991258
; Patent No. US20020141975A1
; GENERAL INFORMATION:
;  APPLICANT: Olmsted, Robert
;  APPLICANT:  Keith, Paula
;  APPLICANT:  Dryga, Sergey
;  APPLICANT:  Caley, Ian
;  APPLICANT:  Maughan, Maureen
;  APPLICANT:  Johnston, Robert
;  APPLICANT:  Davis, Nancy
;  APPLICANT:  Swanstrom, Ronald
;  TITLE OF INVENTION: ALPHAVIRUS VECTORS AND VIROSOMES WITH MODIFIED HIV GENES FOR USE
;  TITLE OF INVENTION:  VACCINES
;  FILE REFERENCE: 01113.0001U3
;  CURRENT APPLICATION NUMBER: US/09/991,258
;  CURRENT FILING DATE:  2001-11-16
;  PRIOR APPLICATION NUMBER: 09/902,537
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: 60/216,995
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 13584
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence; No. US20020141975A1e =
;   OTHER INFORMATION: synthetic construct
US-09-991-258-17

  Query Match           42.3%;  Score 2113.2;  DB 10;  Length 13584;
  Best Local Similarity  64.5%;  Pred. No. 0;
  Matches 3216;  Conservative    0;  Mismatches 1748;  Indels   24;  Gaps    3;

Qy         10 ACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTC 69
              | | | | | |||| ||||| | | ||||||||||| | || |  |||||||  |   ||
Db         30 AATTACCTACCCAAAATGGAGAAAGTTCACGTTGACATCGAGGAAGACAGCCCATTCCTC 89

Qy         70 AAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAAT 129
              |  | || |||||||  || || || ||||||  | |||||  | |||||||||||| |||
Db         90 AGAGCTTTGCAGCGGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCACTGATAAT 149

Qy        130 GACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTC 189
              |||||||| |||||||||||||||||||||||| ||||  |||| || || || |  || ||
Db        150 GACCATGCTAATGCCAGAGCGTTTTCGCATCTGGCTTCAAAACTGATCGAAACGGAGGTG 209

Qy        190 GACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAAT 249
              ||||    | |    ||| | ||||||||||||||||||| | |         |||| ||
Db        210 GACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCCGCCCGCAGAATGTATTCTAAG 269

Qy        250 CACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGG 309
              |||   |||||||||||||| || |||| | | || ||||| |||||||||  |  |   |
Db        270 CACAAGTATCATTGTATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGATTGTATAAG 329

Qy        310 TATGCAGAAAGACTTAAGAAAAGT------GACATTACCGACAAGAACATAGCCTCTAAG 363
              ||||||    |  || |||||||          || || || || ||| |  | | |   ||
Db        330 TATGCAACTAAGCTGAAGAAAAACTGTAAGGAAATAACTGATAAGGAATTGGACAAGAAA 389

Qy        364 GCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATG 423
                |   || ||    | ||||||       || |||   || |||    || ||||  |
Db        390 ATGAAGGAGCTCGCCGCCGTCATGAGCGACCCTGACCTGGAAACTGAGACTATGTGCCTC 449

Qy        424 CACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCA 483
              |||   ||||   ||||| | |||   ||    || || || ||||| ||||| ||||||
Db        450 CACGACGACGAGTCGTGTCGCTACGAAGGGCAAGTCGCTGTTTACCAGGATGTATACGCG 509

Qy        484 GTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGG 543
              ||  | | |||||||    |||| ||||| ||   ||| ||||||||       ||||||
Db        510 GTTGACGGACCGACAAGTCTCTATCACCAAGCCAATAAGGGAGTTAGAGTCGCCTACTGG 569

Qy        544 ATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTAC 603
              |||||||||||||| ||||||||||||||  || ||| |||| ||  | || ||  | |||
Db        570 ATAGGCTTTGACACCACCCCTTTTATGTTTAAGAACTTGGCTGGAGCATATCCATCATAC 629

Qy        604 AACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCA 663
                 || |||||||| ||||| |  || ||    || |||||||| |||||  | | |||
Db        630 TCTACCAACTGGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCCTATGCAGCTCT 689

Qy        664 GATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCT 723
              ||  ||  |||| |    | | ||    | || || |||| ||||||   ·|  ||||
Db        690 GACGTTATGGAGCGGTCACGTAGAGGGATGTCCATTCTTAGAAAGAAGTATTTGAAACCA 749

Qy        724 ACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTA 783
              | || ||  | ||||||| |||| || || ||||||   ||  | ||    |  |
Db        750 TCCAACAATGTTCTATTCTCTGTTGGCTCGACCATCTACCACGAGAAGAGGGACTTACTG 809
```

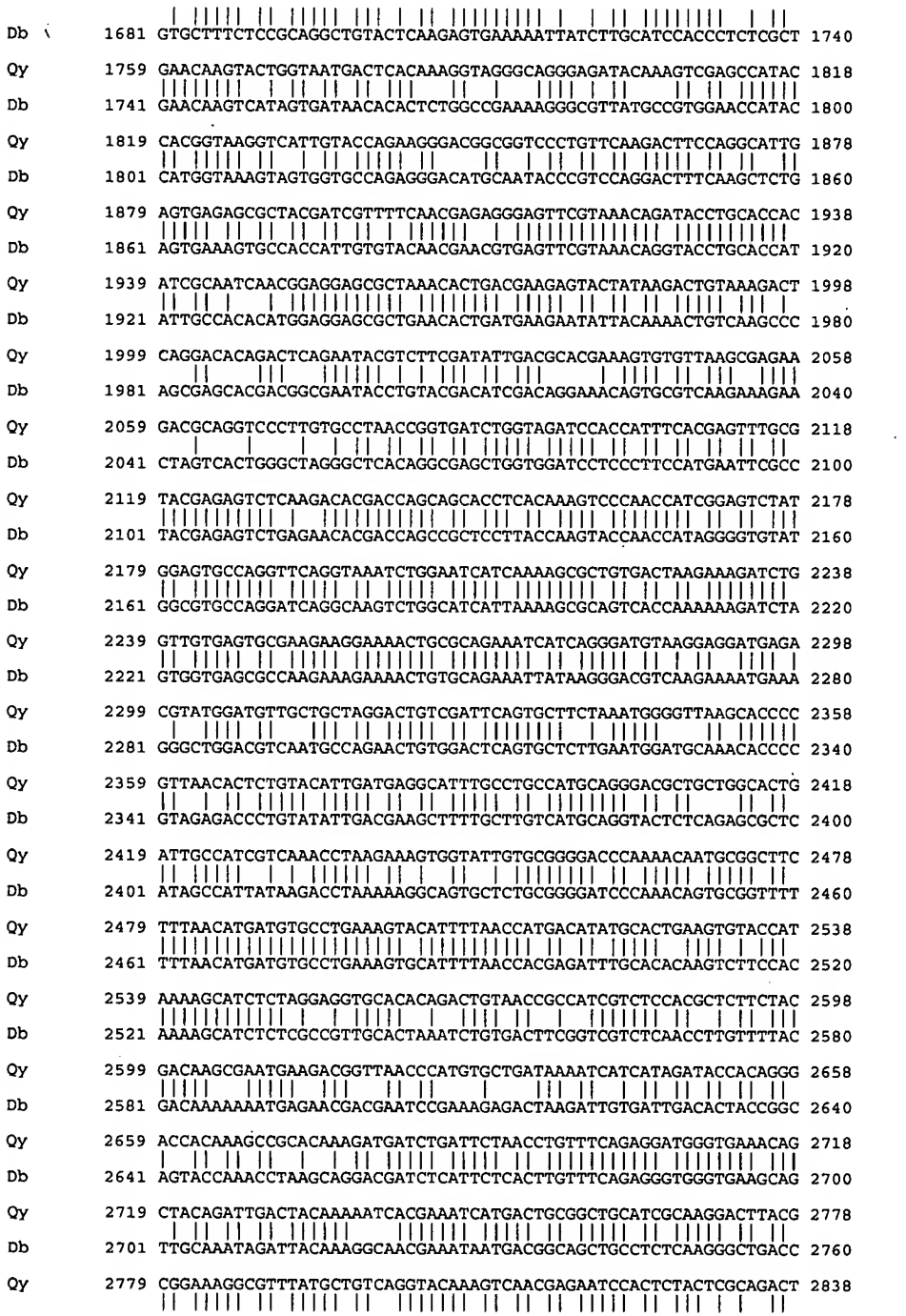


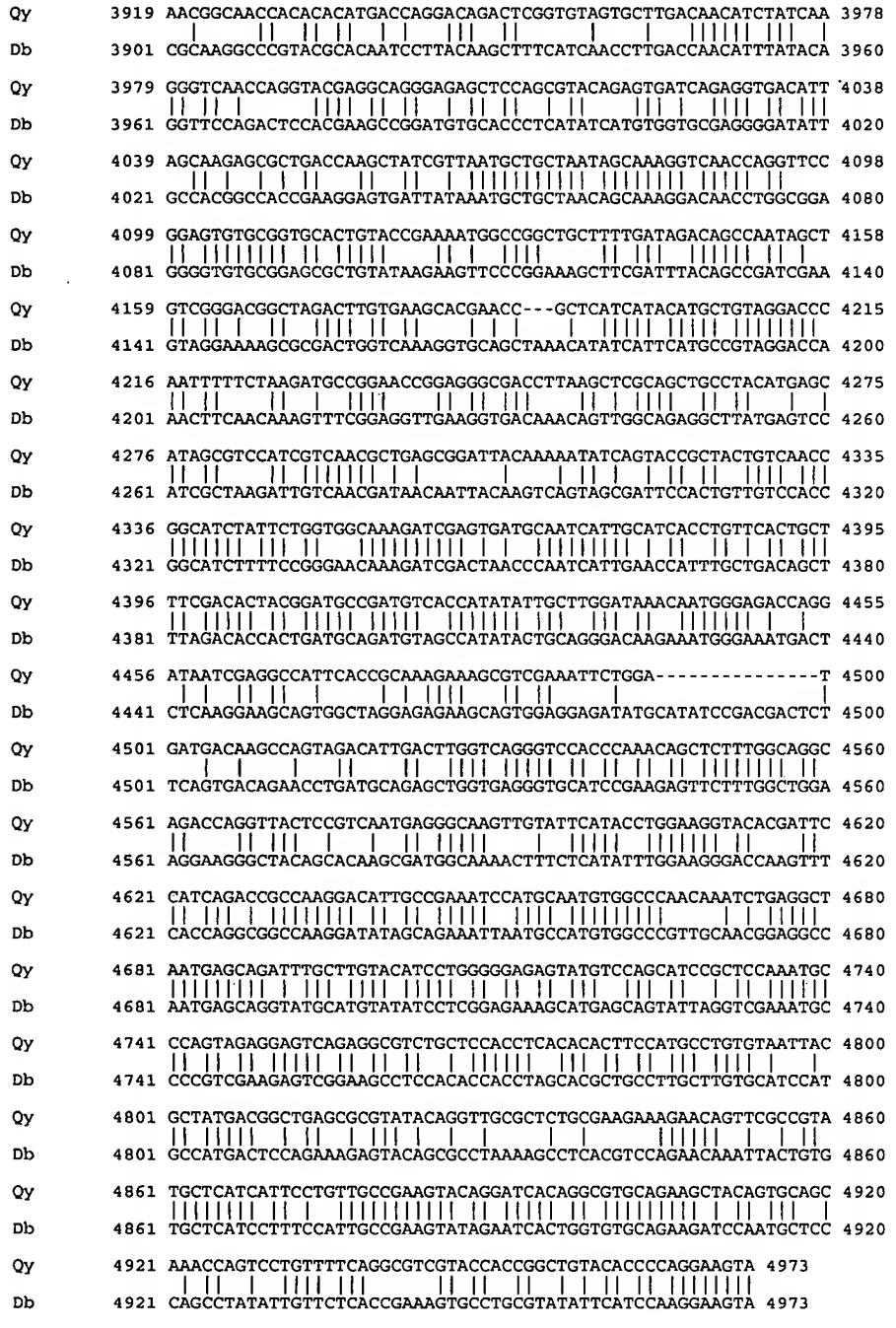




RESULT 5
US-09-991-258-2
; Sequence 2, Application US/09991258
; Patent No. US20020141975A1
; GENERAL INFORMATION:
; APPLICANT: Olmsted, Robert

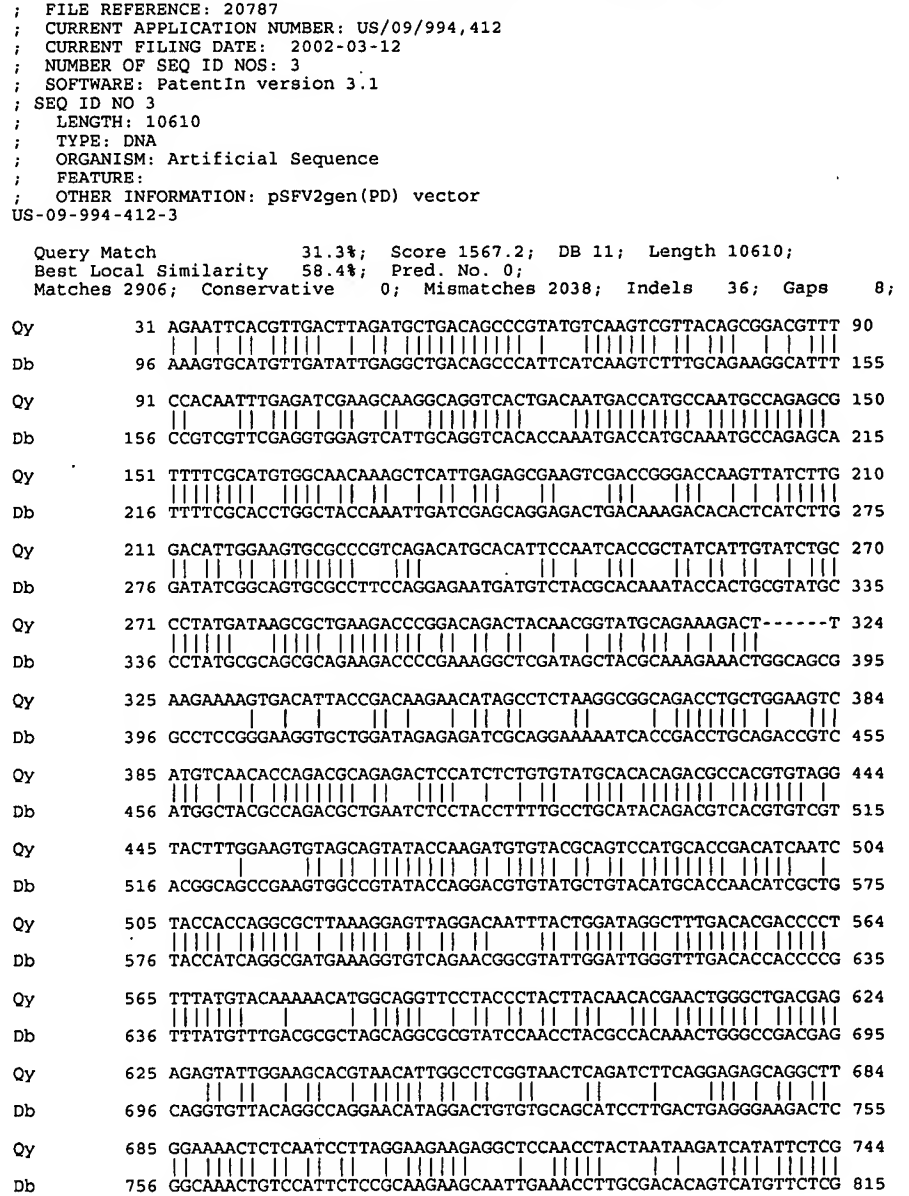
```
;  APPLICANT:  Keith, Paula
;  APPLICANT:  Dryga, Sergey
;  APPLICANT:  Caley, Ian
;  APPLICANT:  Maughan, Maureen
;  APPLICANT:  Johnston, Robert
;  APPLICANT:  Davis, Nancy
;  APPLICANT:  Swanstrom, Ronald
;  TITLE OF INVENTION: ALPHAVIRUS VECTORS AND VIROSOMES WITH MODIFIED HIV GENES FOR USE
;  TITLE OF INVENTION:  VACCINES
;  FILE REFERENCE: 01113.0001U3
;  CURRENT APPLICATION NUMBER: US/09/991,258
;  CURRENT FILING DATE:  2001-11-16
;  PRIOR APPLICATION NUMBER: 09/902,537
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: 60/216,995
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 7479
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence; No. US20020141975A1e =
;   OTHER INFORMATION: synthetic construct
;   NAME/KEY: CDS
;   LOCATION: (1)...(7479)
US-09-991-258-2

  Query Match             42.2%;  Score 2111;  DB 10;  Length 7479;
  Best Local Similarity   64.5%;  Pred. No. 0;
  Matches 3209;  Conservative    0;  Mismatches 1740;  Indels   24;  Gaps    3;

Qy         25 ATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGTCGTTACAGCGG 84
              ||||| | | ||||||||||| | || |  |||||||| |   |||   | || ||||||
Db          1 ATGGAGAAAGTTCACGTTGACATCGAGGAAGACAGCCCATTCCTCAGAGCTTTGCAGCGG 60

Qy         85 ACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACCATGCCAATGCC 144
              |  || || || |||||| | ||||| | |||||||||||| |||||||||| ||||||
Db         61 AGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCACTGATAATGACCATGCTAATGCC 120

Qy        145 AGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACCGGGACCAAGTT 204
              ||||||||||||||| ||||  |||| || || || |  || || ||||    | |
Db        121 AGAGCGTTTTCGCATCTGGCTTCAAAACTGATCGAAACGGAGGTGGACCCATCCGACACG 180

Qy        205 ATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACCGCTATCATTGT 264
              ||| | ||||||||||||||||||| | |         |||| || |||   |||||||||
Db        181 ATCCTTGACATTGGAAGTGCGCCCGCCCGCAGAATGTATTCTAAGCACAAGTATCATTGT 240

Qy        265 ATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGGTATGCAGAAAGACTT 324
              ||||| || |||| | | || ||||| |||||||| |  |   |||||||   | ||
Db        241 ATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGATTGTATAAGTATGCAACTAAGCTG 300

Qy        325 AAGAAA------AGTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACCTGCTG 378
              ||||||      |  || || || || ||| |  | | |   ||   |  || ||
Db        301 AAGAAAAAACTGTAAGGAAATAACTGATAAGGAATTGGACAAGAAAATGAAGGAGCTCGCC 360

Qy        379 GAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACGCCACG 438
              |  ||||||       || |||   || |||    || ||||  | |||    ||
Db        361 GCCGTCATGAGCGACCCTGACCTGGAAACTGAGACTATGTGCCTCCACGACGACGAGTCG 420

Qy        439 TGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCACCGACA 498
              ||| | |||   ||   || || || ||||| ||||| ||||| ||  | | |||||||
Db        421 TGTCGCTACGAAGGGCAAGTCGCTGTTTACCAGGATGTATACGCGGTTGACGGACCGACA 480

Qy        499 TCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTGACACG 558
                 |||| ||||| ||   ||| ||||||||       ||||||||||||||||||||
Db        481 AGTCTCTATCACCAAGCCAATAAGGGAGTTAGAGTCGCCTACTGGATAGGCTTTGACACC 540

Qy        559 ACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACTGGGCT 618
              ||||||||||||  || ||| |||| ||  | || ||  | |||   || ||||||||
Db        541 ACCCCTTTTATGTTTAAGAACTTGGCTGGAGCATATCCATCATACTCTACCAACTGGGCC 600

Qy        619 GACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGGAGAGC 678
              ||||| |  || ||     || |||||||| |||||  | | ||| ||  ||  |||| |
Db        601 GACGAAACCGTGTTAACGGCTCGTAACATAGGCCTATGCAGCTCTGACGTTATGGAGCGG 660

Qy        679 AGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGATCATA 738
              ·| | ||    | || || ||||| ||||||     |  ||||  | || ||  |  ||
Db        661 TCACGTAGAGGGATGTCCATTCTTAGAAAGAAGTATTTGAAACCATCCAACAATGTTCTA 720

Qy        739 TTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGCATCTT 798
              ||||| ||||| || || ||||||   ||  |·||      |  |  | ||||||| ||
Db        721 TTCTCTGTTGGCTCGACCATCTACCACGAGAAGAGGGACTTACTGAGGAGCTGGCACCTG 780

Qy        799 CCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTGTC 858
              ||     || || |||||     || |||    || | |||| || |  |||| ||| || ||
Db        781 CCGTCTGTATTTCACTTACGTGGCAAGCAAAATTACACATGTCGGTGTGAGACTATAGTT 840

Qy        859 AGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAGTT 918
              || || || ||||||||| | ||||  ||| | ||||| |||||  || || || ||   |
Db        841 AGTTGCGACGGGTACGTCGTTAAAAGAATAGCTATCAGTCCAGGCCTGTATGGGAAGCCT 900

Qy        919 GAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATACG 978
                 ||  ||  | || ||||| |||||||| |||||| | |||||||| ||||| ||
Db        901 TCAGGCTATGCTGCTACGATGCACCGCGAGGGATTCTTGTGCTGCAAAGTGACAGACACA 960

Qy        979 CTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCGAT 1038
              |  ||| |||||||| |||||| | ||||| |||||||| |||| | || ||
Db        961 TTAAACGGGGAGAGGGTCTCTTTTCCCGTGTGCACGTATGTGCCAGCTACATTGTGTGAC 1020

Qy       1039 CAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTATTGGTT 1098
              || ||||| || || |||||||| || || |||| ||| ||||| |||||||| |||||
Db       1021 CAAATGACTGGCATACTGGCAACAGATGTCAGTGCGGACGACGCGCAAAAACTGCTGGTT 1080

Qy       1099 GGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAATGCAG 1158
              ||||||||||  | || |||||||| ||| | || || ||||| || || || ||| |
Db       1081 GGGCTCAACCAGCGTATAGTCGTCAACGGTCGCACCCAGAGAAACACCAATACCATGAAA 1140

Qy       1159 AACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATCGTGCC 1218
              || || || || || || || || |||||||| ||| | ||||||||    ||| ||  |
Db       1141 AATTACCTTTTGCCCGTAGTGGCCCAGGCATTTGCTAGGTGGGCAAAGGAATATAAGGAA 1200

Qy       1219 GACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCTGCTGC 1278
              ||    || || || |    ||||||  | || ||  |    |    ||||| || ||
Db       1201 GATCAAGAAGATGAAAGGCCACTAGGACTACGAGATAGACAGTTAGTCATGGGGTGTTGT 1260

Qy       1279 TGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCTGGTACGCAAACAATT 1338
              |||||||| |  |  || || || ||||| || || |||   || | ||| ||||| ||
Db       1261 TGGGCTTTTAGAAGGCACAAGATAACATCTATTTATAAGCGCCCGGATACCCAAACCATC 1320

Qy       1339 AAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCACGGGCTC 1398
              |  |||||       ||  || ||||||| ||| | || |  |  |||  ||    |
Db       1321 ATCAAAGTGAACAGCGATTTCCACTCATTCGTGCTGCCCAGGATAGGCAGTAACACATTG 1380

Qy       1399 GATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACTGTCAAACCCGCACCG 1458
              || || ||  |  |   ||  ||| |    || | ||          | ||  ||||
Db       1381 GAGATCGGGCTGAGAACAAGAATCAGGAAAATGTTAGAGGAGCACAAGGAGCCGTCACCT 1440

Qy       1459 GCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAGAAGTG 1518
                |||||    |  || ||  |  |     ||  |    | || ||| | || |||
Db       1441 CTCATTACCGCCGAGGACGTACAAGAAGCTAAGTGCGCAGCCGATGAGGCTAAGGAGGTG 1500

Qy       1519 GCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAAAAGAG 1578
              || ||| || ||| |  | | ||| || ||||| |||     ||||  | ||  |
Db       1501 CGTGAAGCCGAGGAGTTGCGCGCAGCTCTACCACCTTTGGCAGCTGATGTTGAGGAGCCC 1560

Qy       1579 ACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGAGCAGGTAGCGTGGAGACA 1638
              ||  | || || || || ||| | ||  | |||||||| || || ||    |||||||||
Db       1561 ACTCTGGAAGCCGATGTCGACTTGATGTTACAAGAGGCTGGGGCCGGCTCAGTGGAGACA 1620

Qy       1639 CCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAGATTGGGTCTTACGCT 1698
              || || ||    || | ||| || || ||| | ||||| || ||||| || |||||||||
Db       1621 CCTCGTGGCTTGATAAAGGTTACCAGCTACGCTGGCGAGGACAAGATCGGCTCTTACGCT 1680

Qy       1699 ATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGTATCCACCCATTGGCG 1758
```

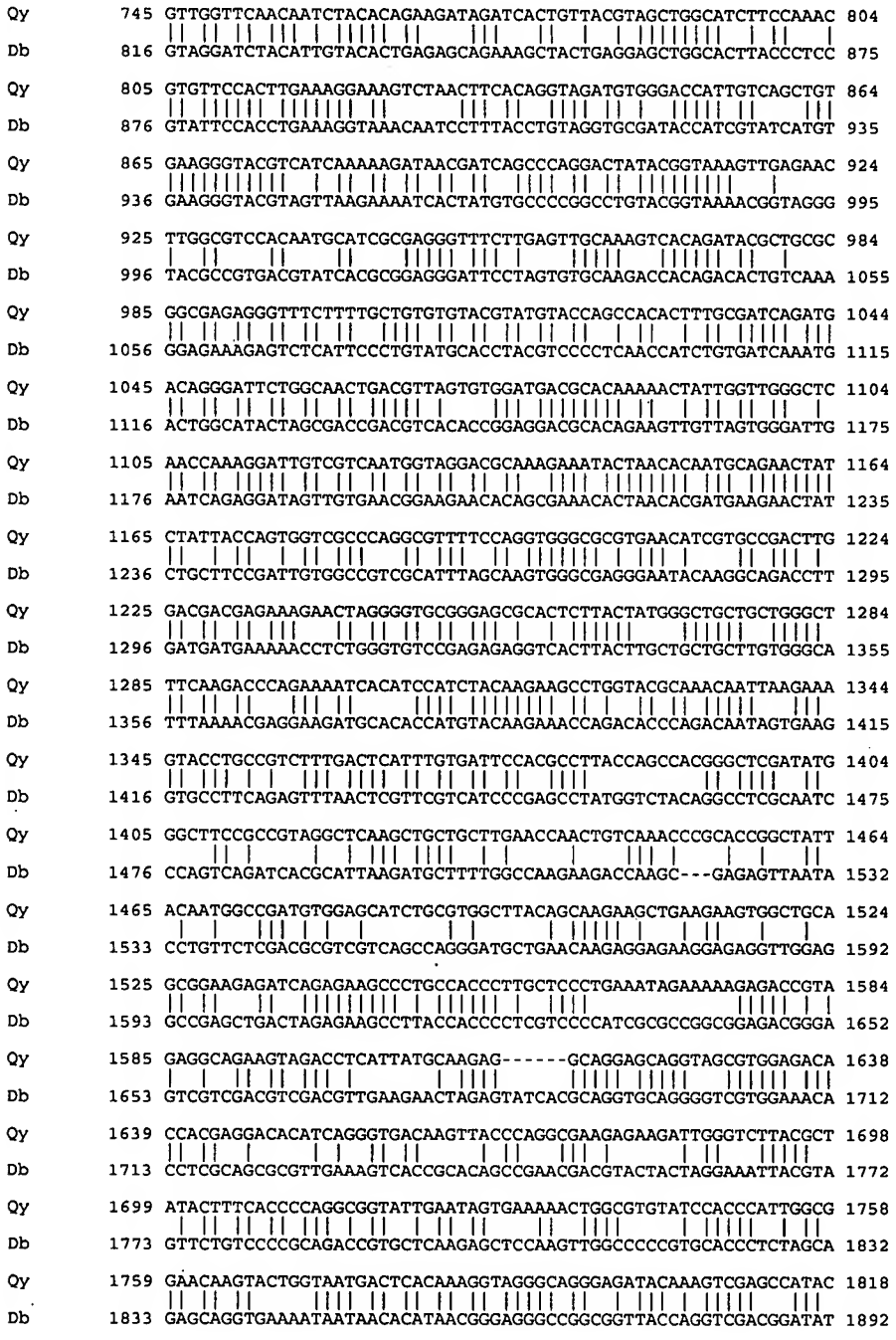

Db 1681 GTGCTTTCTCCGCAGGCTGTACTCAAGAGTGAAAAATTATCTTGCATCCACCCTCTCGCT 1740
Qy 1759 GAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATACAAAGTCGAGCCATAC 1818
Db 1741 GAACAAGTCATAGTGATAACACACTCTGGCCGAAAAGGGCGTTATGCCGTGGAACCATAC 1800
Qy 1819 CACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAAGACTTCCAGGCATTG 1878
Db 1801 CATGGTAAAGTAGTGGTGCCAGAGGGACATGCAATACCCGTCCAGGACTTTCAAGCTCTG 1860
Qy 1879 AGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAACAGATACCTGCACCAC 1938
Db 1861 AGTGAAAGTGCCACCATTGTGTACAACGAACGTGAGTTCGTAAACAGGTACCTGCACCAT 1920
Qy 1939 ATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTATAAGACTGTAAAGACT 1998
Db 1921 ATTGCCACACATGGAGGAGCGCTGAACACTGATGAAGAATATTACAAAACTGTCAAGCCC 1980
Qy 1999 CAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAGTGTGTTAAGCGAGAA 2058
Db 1981 AGCGAGCACGACGGCGAATACCTGTACGACATCGACAGGAAACAGTGCGTCAAGAAAGAA 2040
Qy 2059 GACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCATTTCACGAGTTTGCG 2118
Db 2041 CTAGTCACTGGGCTAGGGCTCACAGGCGAGCTGGTGGATCCTCCCTTCCATGAATTCGCC 2100
Qy 2119 TACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCAACCATCGGAGTCTAT 2178
Db 2101 TACGAGAGTCTGAGAACACGACCAGCCGCTCCTTACCAAGTACCAACCATAGGGGTGTAT 2160
Qy 2179 GGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTGACTAAGAAAGATCTG 2238
Db 2161 GGCGTGCCAGGATCAGGCAAGTCTGGCATCATTAAAAGCGCAGTCACCAAAAAAGATCTA 2220
Qy 2239 GTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGATGTAAGGAGGATGAGA 2298
Db 2221 GTGGTGAGCGCCAAGAAAGAAAACTGTGCAGAAATTATAAGGGACGTCAAGAAAATGAAA 2280
Qy 2299 CGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAATGGGGTTAAGCACCCC 2358
Db 2281 GGGCTGGACGTCAATGCCAGAACTGTGGACTCAGTGCTCTTGAATGGATGCAAACACCCC 2340
Qy 2359 GTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGGACGCTGCTGGCACTG 2418
Db 2341 GTAGAGACCCTGTATATTGACGAAGCTTTTGCTTGTCATGCAGGTACTCTCAGAGCGCTC 2400
Qy 2419 ATTGCCATCGTCAAACCTAAGAAAGTGGTATTGTGCGGGGACCCAAAACAATGCGGCTTC 2478
Db 2401 ATAGCCATTATAAGACCTAAAAAGGCAGTGCTCTGCGGGGATCCCAAACAGTGCGGTTTT 2460
Qy 2479 TTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGCACTGAAGTGTACCAT 2538
Db 2461 TTTAACATGATGTGCCTGAAAGTGCATTTTAACCACGAGATTTGCACACAAGTCTTCCAC 2520
Qy 2539 AAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATCGTCTCCACGCTCTTCTAC 2598
Db 2521 AAAAGCATCTCTCGCCGTTGCACTAAATCTGTGACTTCGGTCGTCTCAACCTTGTTTTAC 2580
Qy 2599 GACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATCATCATAGATACCACAGGG 2658
Db 2581 GACAAAAAAAATGAGAACGACGAATCCGAAAGAGACTAAGATTGTGATTGACACTACCGGC 2640
Qy 2659 ACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGAGGATGGGTGAAACAG 2718
Db 2641 AGTACCAAACCTAAGCAGGACGATCTCATTCTCACTTGTTTCAGAGGGTGGGTGAAGCAG 2700
Qy 2719 CTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCTGCATCGCAAGGACTTACG 2778
Db 2701 TTGCAAATAGATTACAAAGGCAACGAAATAATGACGGCAGCTGCCTCTCAAGGGCTGACC 2760
Qy 2779 CGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCACTCTACTCGCAGACT 2838
Db 2761 CGTAAAGGTGTGTATGCCGTTCGGTACAAGGTGAATGAAAATCCTCTGTACGCACCCACC 2820
Qy 2839 TCTGAGCACGTGAACGTGTTACTTACACGCACAGAAAAACGCATTGTCTGGAAGACGCTA 2898
Db 2821 TCAGAACATGTGAACGTCCTACTGACCCGCACGGAGGACCGCATCGTGTGGAAAACACTA 2880
Qy 2899 GCTGGTGATCCCTGGATAAAGACACTTACAGCTAAATATCCCGGGGATTTCACGGCTTCA 2958
Db 2881 GCCGGCGACCCATGGATAAAAACACTGACTGCCAAGTACCCTGGGAATTTCACTGCCACG 2940
Qy 2959 TTGGACGACTGGCAGCGCGAACACGACGCCATTATGGCACGCGTTCTTGATAAGCCGCAG 3018
Db 2941 ATAGAGGAGTGGCAAGCAGAGCATGATGCCATCATGAGGCACATCTTGGAGAGACCGGAC 3000
Qy 3019 ACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAGGCTTTAGAGCCAGTC 3078
Db 3001 CCTACCGACGTCTTCCAGAATAAGGCAAACGTGTGTTGGGCCAAGGCTTTAGTGCCGGTG 3060
Qy 3079 TTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAGACGTTGCACCCATTCAAG 3138
Db 3061 CTGAAGACCGCTGGCATAGACATGACCACTGAACAATGGAACACTGTGGATTATTTTGAA 3120
Qy 3139 CATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTTTGCACCAGGTTCTTTGGA 3198
Db 3121 ACGGACAAAGCTCACTCAGCAGAGATAGTATTGAACCAACTATGCGTGAGGTTCTTTGGA 3180
Qy 3199 GTAGACCTGGACAGTGGGTTATTTTCCGCTCCTACCGTCGCACTTACTTACAGGGATCAG 3258
Db 3181 CTCGATCTGGACTCCGGTCTATTTTCTGCACCCACTGTTCCGTTATCCATTAGGAATAAT 3240
Qy 3259 CACTGGGATAACTCGCCAGGGAAGAACATGTATGGGCTTAATAGAGAGGTAGCAAAGGAG 3318
Db 3241 CACTGGGATAACTCCCCGTCGCCTAACATGTACGGGCTGAATAAAGAAGTGGTCCGTCAG 3300
Qy 3319 TTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACAGGCAGGGTAGCTGATATA 3378
Db 3301 CTCTCTCGCAGGTACCCACAACTGCCTCGGGCAGTTGCCACTGGAAGAGTCTATGACATG 3360
Qy 3379 AGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTGGTTCCATTAAATCGCCGG 3438
Db 3361 AACACTGGTACACTGCGCAATTATGATCCGCGCATAAACCTAGTACCTGTAAACAGAAGA 3420
Qy 3439 TTGCCCCACTCGTTGATCGTTGACCACAAAGGACAGGGTACAACTGATCACAGCGGATTC 3498
Db 3421 CTGCCTCATGCTTTAGTCCTCCACCATAATGAACACCCACAGAGTGACTTTTCTTCATTC 3480
Qy 3499 CTATCTAAGATGAAGGGCAAATCTGTGTTGGTGATCGGCGATCCTATCAGCATTCCAGGG 3558
Db 3481 GTCAGCAAATTGAAGGGCAGAACTGTCCTGGTGGTCGGGGAAAAGTTGTCCGTCCCAGGC 3540
Qy 3559 AAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATACCATCAGGTGTGATCTCGATTTG 3618
Db 3541 AAAATGGTTGACTGGTTGTCAGACCGGCCTGAGGCTACCTTCAGAGCTCGGCTGGATTTA 3600
Qy 3619 GGAATACCTAGCCATGTCGGTAAATATGACATTATCTTTGTCAATGTTAGGACCCCGTAC 3678
Db 3601 GGCATCCCAGGTGATGTGCCCAAATATGACATAATATTTGTTAATGTGAGGACCCCATAT 3660
Qy 3679 AGGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCACCACAGCATGCTAACGTGT 3738
Db 3661 AAATACCATCACTATCAGCAGTGTGAAGACCATGCCATTAAGCTTAGCATGTTGACCAAG 3720
Qy 3739 AAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCTATAGGGTATGGGCTTGCT 3798
Db 3721 AAAGCTTGTCTGCATCTGAATCCCGGCGGAACCTGTGTCAGCATAGGTTATGGTTACGCT 3780
Qy 3799 GATCGCGCAACCGAGAATATCATCACTGCGGTGGCACGCTCATTTAGGTTTACCCGTGTC 3858
Db 3781 GACAGGGCCAGCGAAAGCATCATTGGTGCTATAGCGCGGCAGTTCAAGTTTTCCCGGGTA 3840
Qy 3859 TGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTCGTGTTCTTCGGCAAGGAC 3918
Db 3841 TGCAAACCGAAATCCTCACTTGAAGAGACGGAAGTTCTGTTTGTATTCATTGGGTACGAT 3900

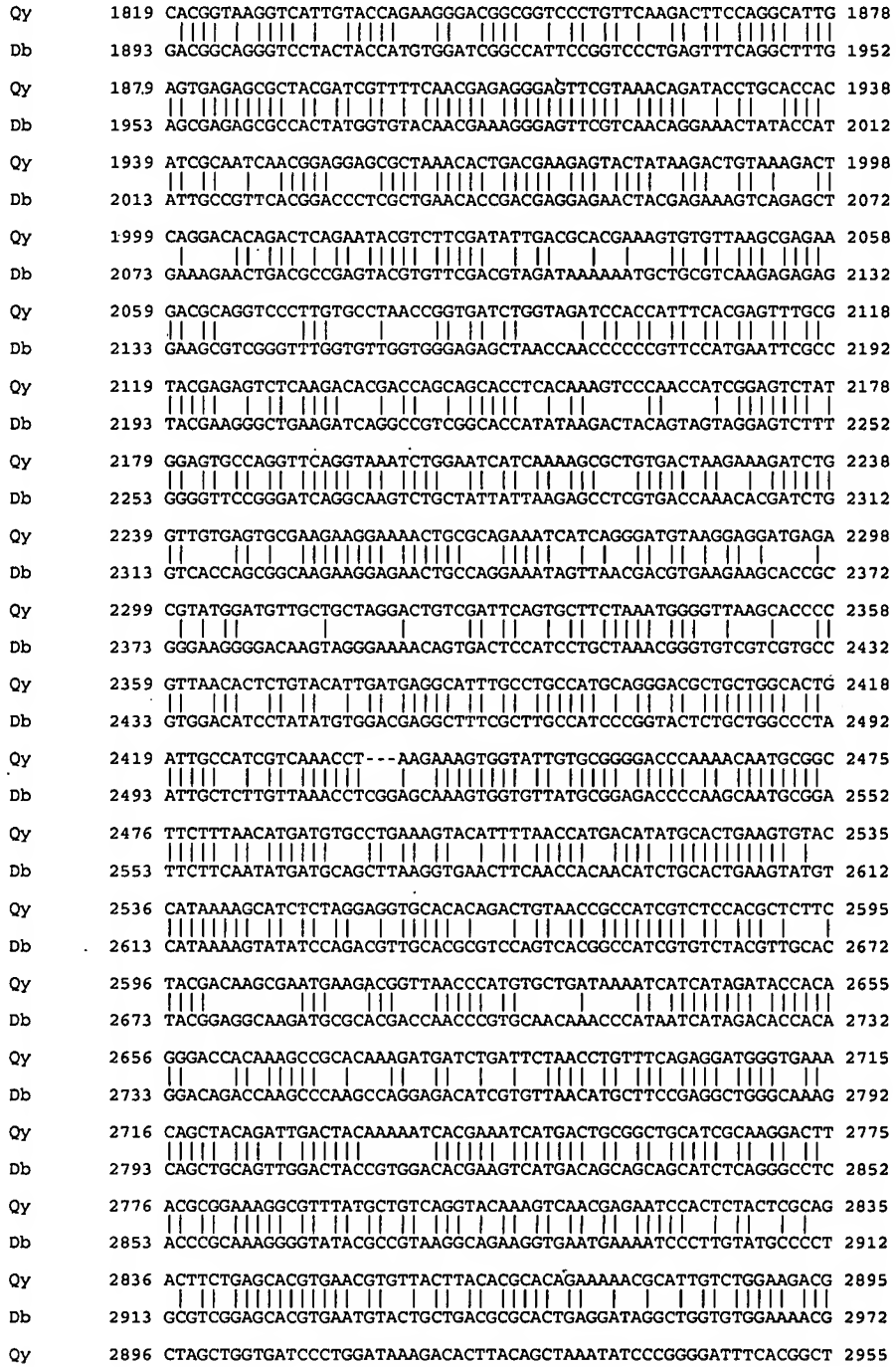


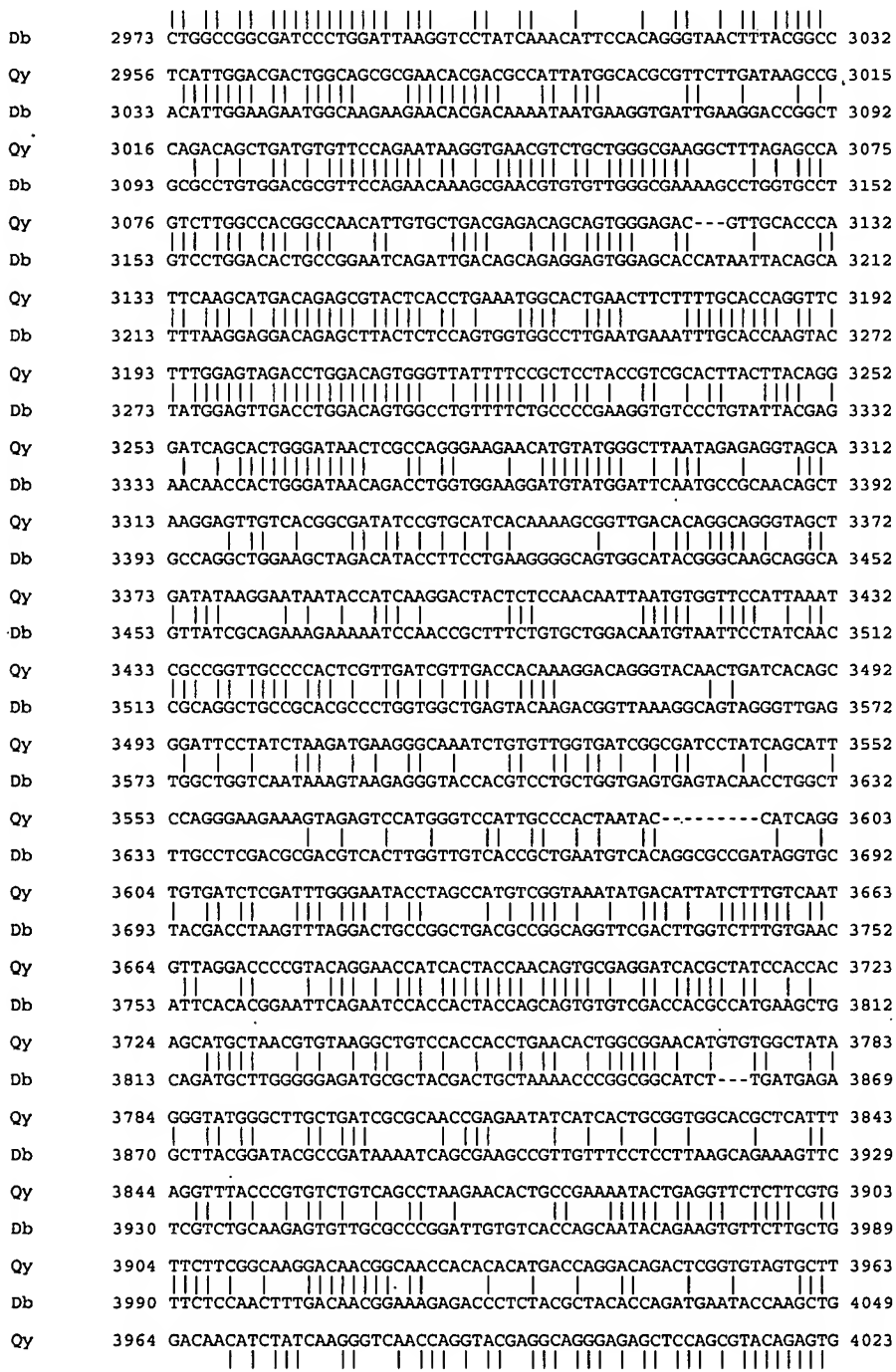
RESULT 6
US-09-994-412-3
; Sequence 3, Application US/09994412
; Publication No. US20030059943A1
; GENERAL INFORMATION:
; APPLICANT: Certa, Ulrich
; APPLICANT: Lundstrom, Kenneth
; TITLE OF INVENTION: Inhibition of Expression of a Target Gene
; FILE REFERENCE: 20787
; CURRENT APPLICATION NUMBER: US/09/994,412
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 10610
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: pSFV2gen(PD) vector
US-09-994-412-3

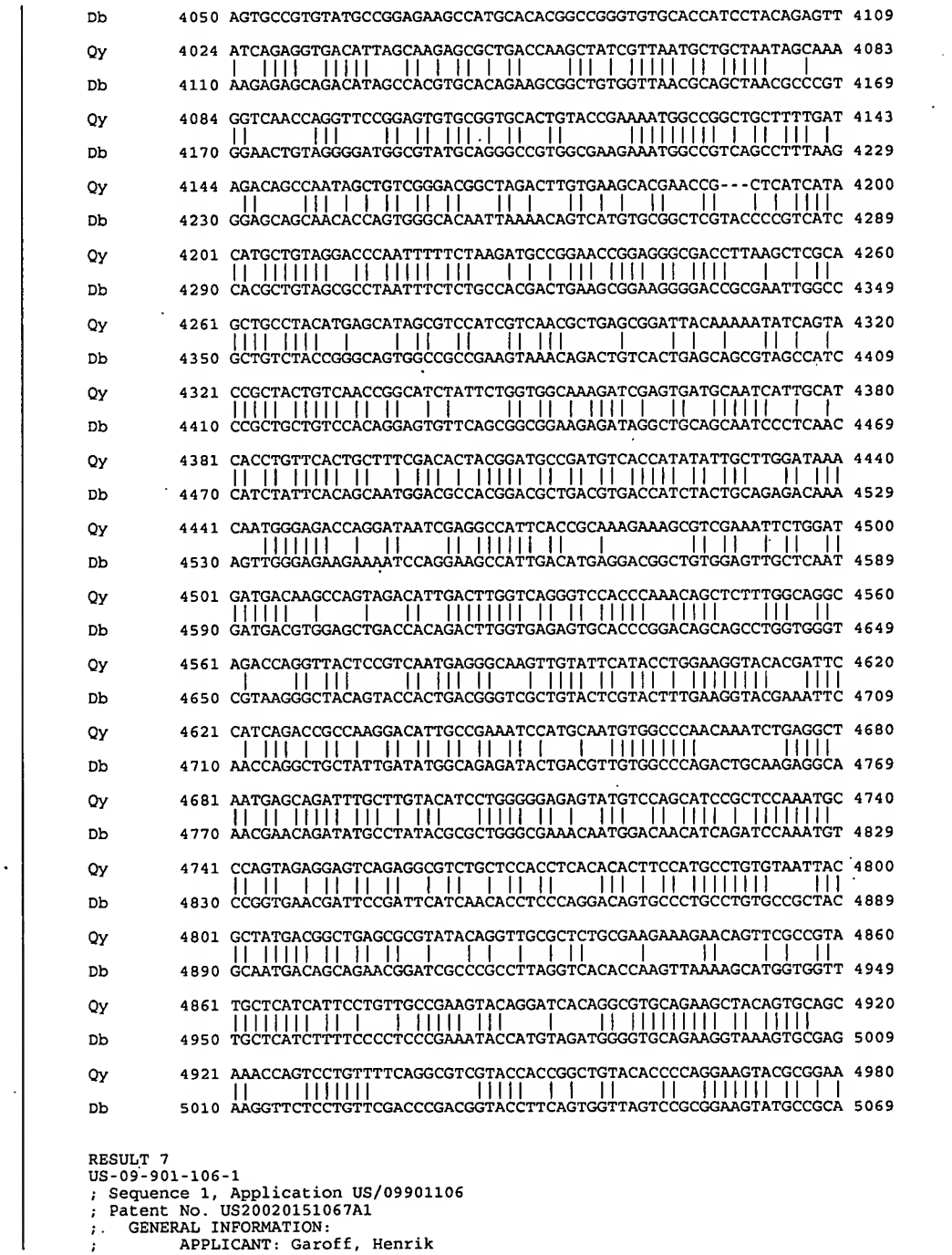
Query Match 31.3%; Score 1567.2; DB 11; Length 10610;
Best Local Similarity 58.4%; Pred. No. 0;
Matches 2906; Conservative 0; Mismatches 2038; Indels 36; Gaps 8;





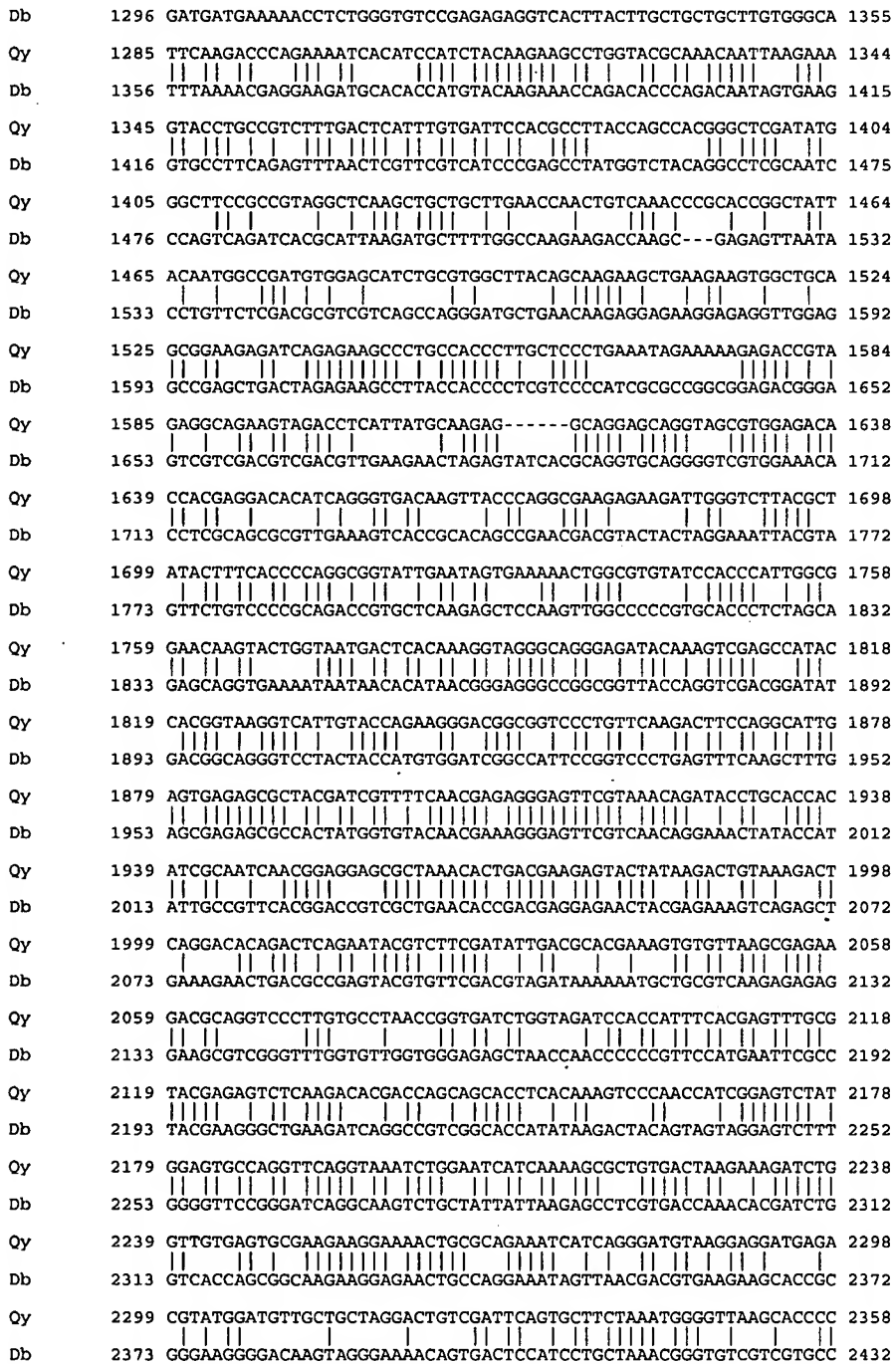


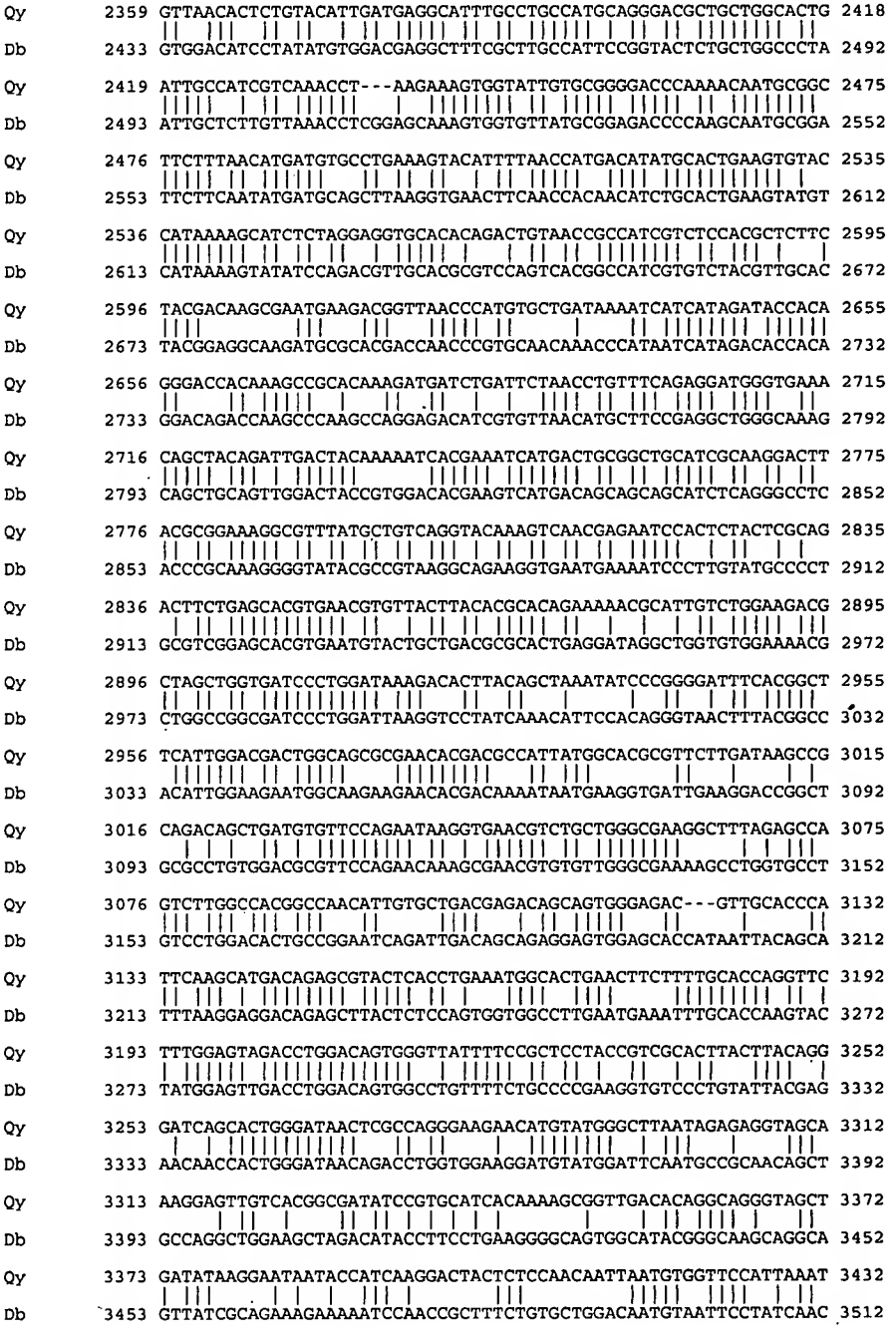


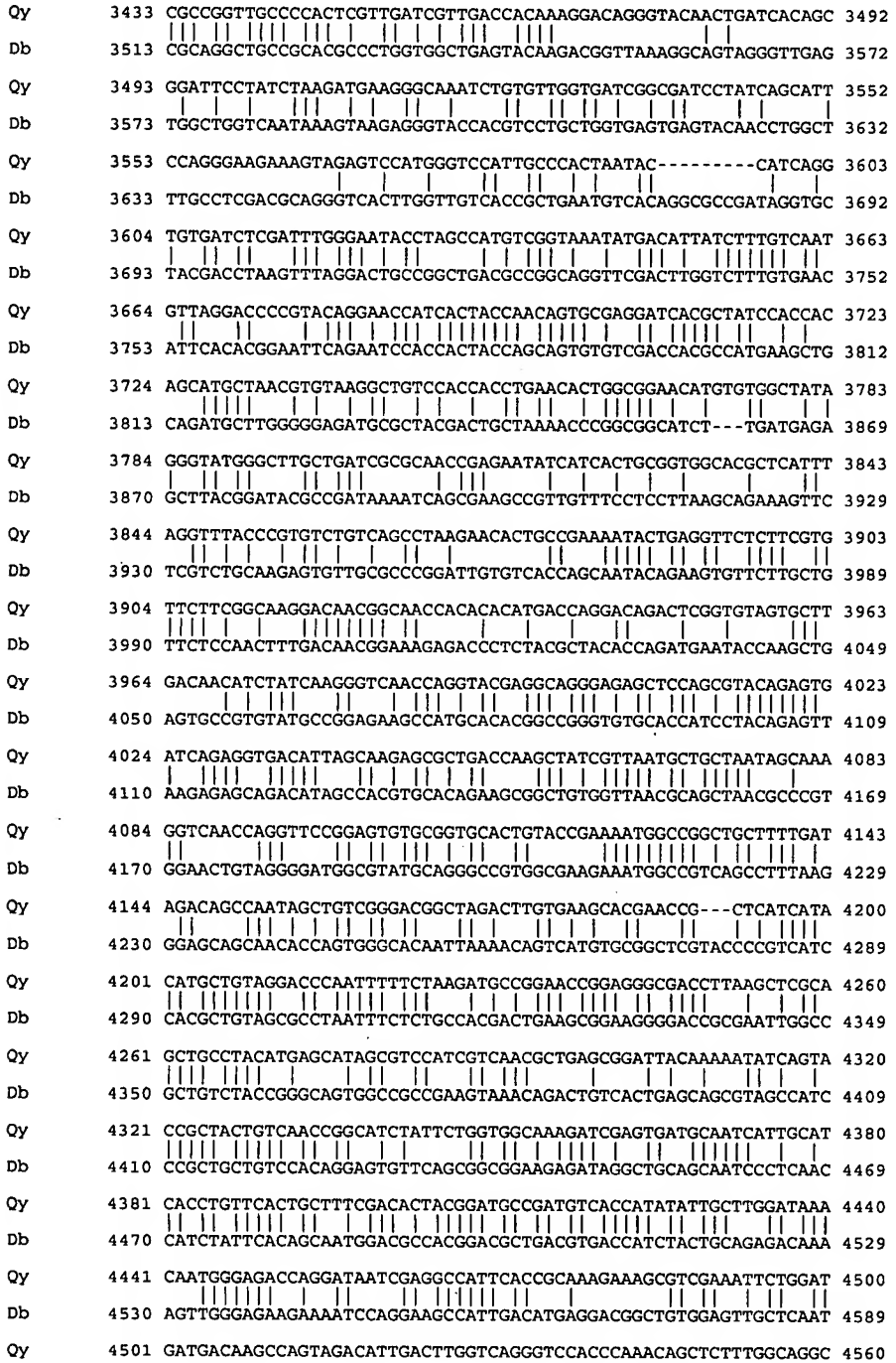


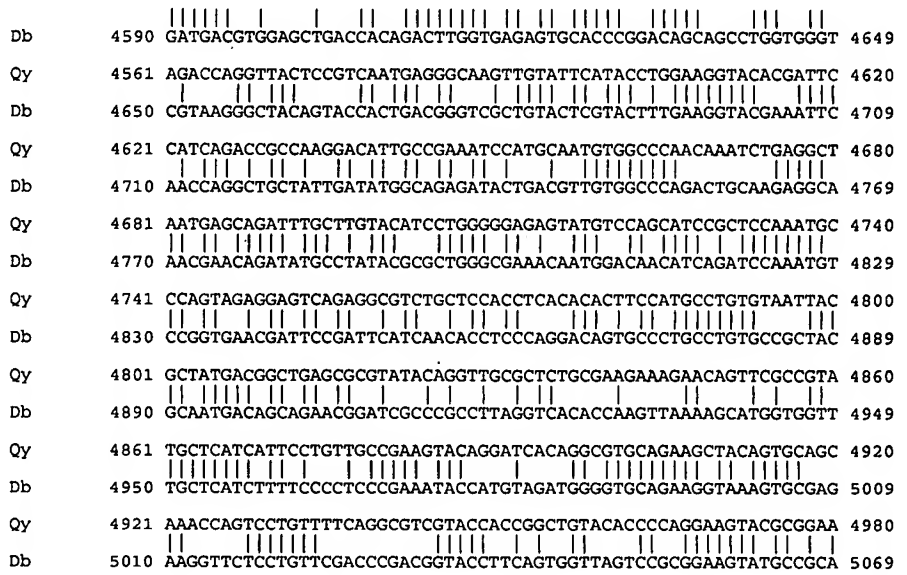
RESULT 7
US-09-901-106-1
; Sequence 1, Application US/09901106
; Patent No. US20020151067A1
;. GENERAL INFORMATION:
; APPLICANT: Garoff, Henrik

```
;                 Liljestrom, Peter
;       TITLE OF INVENTION: DNA Expression Systems Based on
;                           Alphaviruses
;       NUMBER OF SEQUENCES: 27
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Birch, Stewart, Kolasch & Birch
;            STREET: P.O. Box 747
;            CITY: Falls Church
;            STATE: Virginia
;            COUNTRY: USA
;            ZIP: 22040-0747
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/901,106
;            FILING DATE: 10-Jul-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/07/920,281C
;            FILING DATE: 13-AUG-1992
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Murphy Jr., Gerald M.
;            REGISTRATION NUMBER: 28,977
;            REFERENCE/DOCKET NUMBER: 828-103P
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 703-241-1300
;            TELEFAX: 703-241-2848
;            TELEX: 248345
;   INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 11517 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: RNA (genomic)
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       ORIGINAL SOURCE:
;            ORGANISM: Semliki Forest Virus
;       FEATURE:
;            NAME/KEY:  -
;            LOCATION:  1..11517
;            OTHER INFORMATION:  /label= genome
;            /note= "Semliki Forest Virus complete nucleotide
;            sequence, presented as a cloned DNA sequence; see
;            Figure 5."
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  87..7379
;            OTHER INFORMATION:  /product= "SFV polyprotein"
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  7421..11179
;            OTHER INFORMATION:  /product= "SFV polyprotein"
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-901-106-1

  Query Match           31.3%;  Score 1564;  DB 10;  Length 11517;
  Best Local Similarity  58.3%;  Pred. No. 0;
  Matches 2904;  Conservative    0;  Mismatches 2040;  Indels   36;  Gaps    8;

Qy         31 AGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGTCGTTACAGCGGACGTTT 90
              | | | || |||||  | || ||||||||||| |  ||||||| || |||  | | |||
Db         96 AAAGTGCATGTTGATATTGAGGCTGACAGCCCATTCATCAAGTCTTTGCAGAAGGCATTT 155

Qy         91 CCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACCATGCCAATGCCAGAGCG 150
              ||    || ||| | ||  ||  |||||||||    ||||||||||| ||||||||||
Db        156 CCGTCGTTCGAGGTGGAGTCATTGCAGGTCACACCAAATGACCATGCAAATGCCAGAGCA 215

Qy        151 TTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACCGGGACCAAGTTATCTTG 210
              ||||||||  |||| || ||  | || |||   ||    |||  |||  | | ||||||
Db        216 TTTTCGCACCTGGCTACCAAATTGATCGAGCAGGAGACTGACAAAGACACACTCATCTTG 275

Qy        211 GACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACCGCTATCATTGTATCTGC 270
              || || || ||||||||   |||           || |  |||   || || ||  | |||
Db        276 GATATCGGCAGTGCGCCTTCCAGGAGAATGATGTCTACGCACAAATACCACTGCGTATGC 335

Qy        271 CCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGGTATGCAGAAAGACT------T 324
              ||||||   ||||| |||||||| || || ||  |  | || ||| | | |||
Db        336 CCTATGCGCAGCGCAGAAGACCCCGAAAGGCTCGATAGCTACGCAAAGAAACTGGCAGCG 395

Qy        325 AAGAAAAGTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACCTGCTGGAAGTC 384
                  |  |  |   || |  | || ||   ||     | |||||||  |   |||
Db        396 GCCTCCGGGAAGGTGCTGGATAGAGAGATCGCAGGAAAAATCACCGACCTGCAGACCGTC 455

Qy        385 ATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACGCCACGTGTAGG 444
              ||| | || |||||||| ||  |||| |  | ||  |||| |||||||| |||||||  |
Db        456 ATGGCTACGCCAGACGCTGAATCTCCTACCTTTTGCCTGCATACAGACGTCACGTGTCGT 515

Qy        445 TACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCACCGACATCAATC 504
                  |   || || |||||||| || ||||| || || |||||||| |||||  |
Db        516 ACGGCAGCCGAAGTGGCCGTATACCAGGACGTGTATGCTGTACATGCACCAACATCGCTG 575

Qy        505 TACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTGACACGACCCCT 564
              ||||| |||||| | ||||| || || ||    || ||||| || |||||||| |||||
Db        576 TACCATCAGGCGATGAAAGGTGTCAGAACGGCGTATTGGATTGGGTTTGACACCACCCCG 635

Qy        565 TTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACTGGGCTGACGAG 624
              |||||||    |     | ||| || || |||  ||| |||||||| ||||||
Db        636 TTTATGTTTGACGCGCTAGCAGGCGCGTATCCAACCTACGCCACAAACTGGGCCGACGAG 695

Qy        625 AGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGGAGAGCAGGCTT 684
                 || ||  | ||  | ||||| || ||  ||     ||    |   ||| | || ||
Db        696 CAGGTGTTACAGGCCAGGAACATAGGACTGTGTGCAGCATCCTTGACTGAGGGAAGACTC 755

Qy        685 GGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGATCATATTCTCG 744
              || ||||| || || ||  | ||||||    |  |||||    | |   |||| ||||||
Db        756 GGCAAACTGTCCATTCTCCGCAAGAAGCAATTGAAACCTTGCGACACAGTCATGTTCTCG 815

Qy        745 GTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGCATCTTCCAAAC 804
              || || || ||| | ||||| ||    |||  ||  |  | ||||||||  | ||   |
Db        816 GTAGGATCTACATTGTACACTGAGAGCAGAAAGCTACTGAGGAGCTGGCACTTACCCTCC 875

Qy        805 GTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTGTCAGCTGT 864
              || |||||| ||||||| ||      ||| ||  |||| ||  | |||||  ||     |||
Db        876 GTATTCCACCTGAAAGGTAAACAATCCTTTACCTGTAGGTGCGATACCATCGTATCATGT 935

Qy        865 GAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAGTTGAGAAC 924
              ||||||||||  | || || || || ||  |||| || || |||||||||   |
Db        936 GAAGGGTACGTAGTTAAGAAAATCACTATGTGCCCCGGCCTGTACGGTAAAACGGTAGGG 995

Qy        925 TTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATACGCTGCGC 984
              |  ||     ||     ||    ||||| ||| |     |||||   ||||||  ||   |
Db        996 TACGCCGTGACGTATCACGCGGAGGGATTCCTAGTGTGCAAGACCACAGACACTGTCAAA 1055

Qy        985 GGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCGATCAGATG 1044
              || || || || || ||  |||| || || || || ||  | ||  | || ||||| |||
Db       1056 GGAGAAAGAGTCTCATTCCCTGTATGCACCTACGTCCCCTCAACCATCTGTGATCAAATG 1115

Qy       1045 ACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTATTGGTTGGGCTC 1104
              || || || || || || ||||| |    ||| |||||||| ||  | || || ||  |
Db       1116 ACTGGCATACTAGCGACCGACGTCACACCGGAGGACGCACAGAAGTTGTTAGTGGGATTG 1175

Qy       1105 AACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAATGCAGAACTAT 1164
              || || ||||| || || || || || ||  |||| |||||||| ||| ||||||||
Db       1176 AATCAGAGGATAGTTGTGAACGGAAGAACACAGCGAAACACTAACACGATGAAGAACTAT 1235

Qy       1165 CTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATCGTGCCGACTTG 1224
              ||  | ||  | || |||    || |||  || |||||||  | |||  |    || ||| |
Db       1236 CTGCTTCCGATTGTGGCCGTCGCATTTAGCAAGTGGGCGAGGGAATACAAGGCAGACCTT 1295

Qy       1225 GACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCTGCTGCTGGGCT 1284
              || || || |||   || || || || ||| |  | ||||||    ||||||  ||||||
```



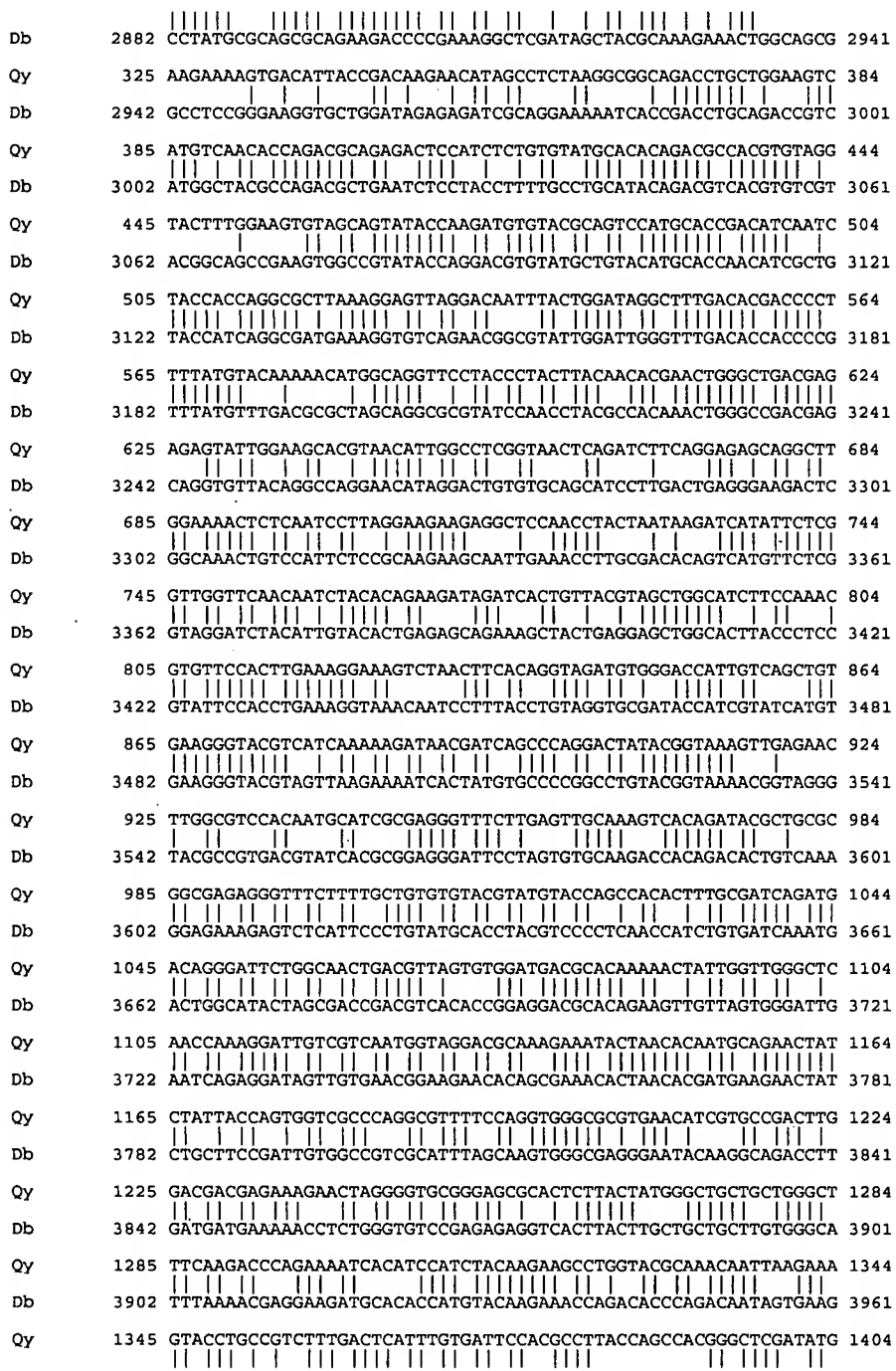


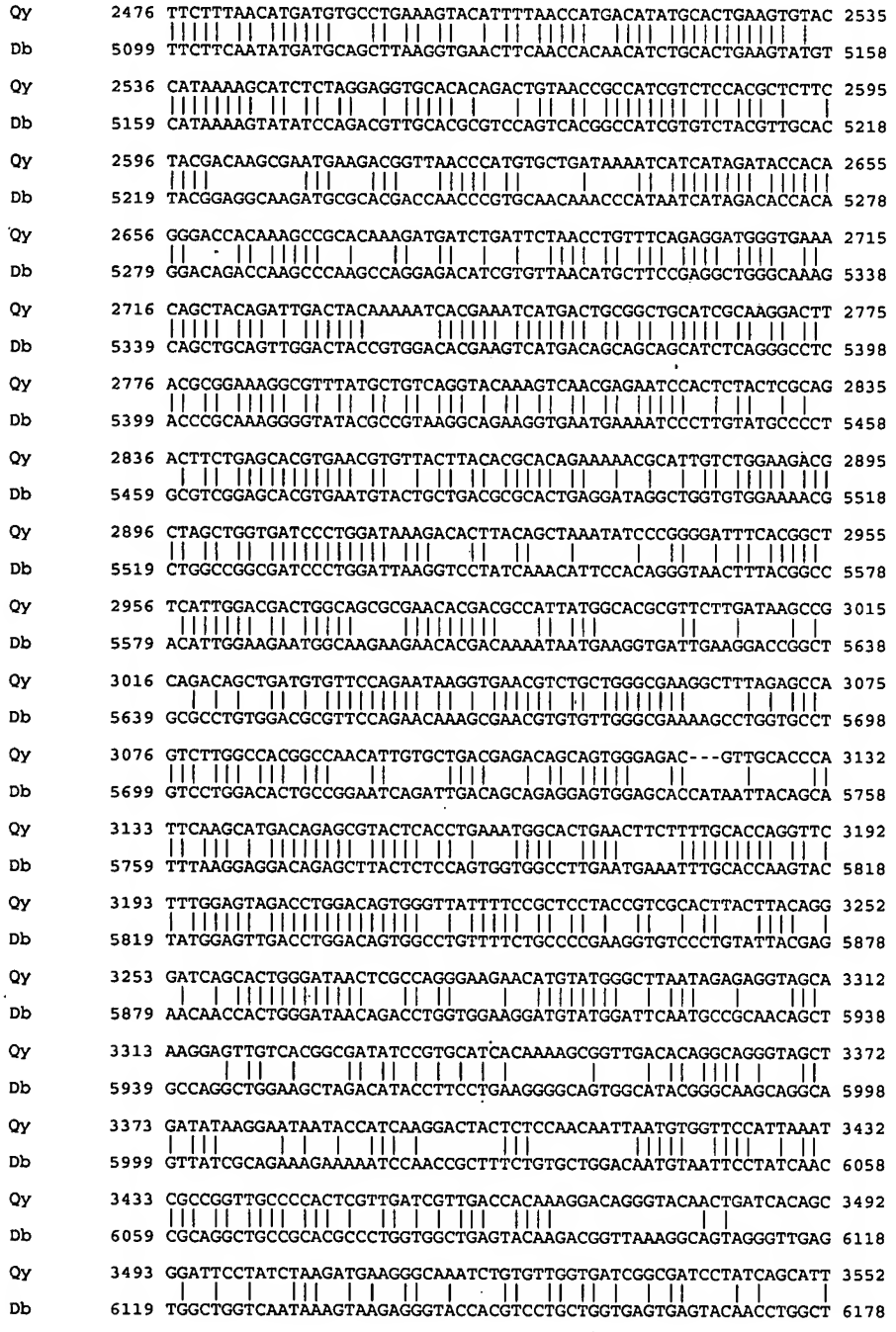




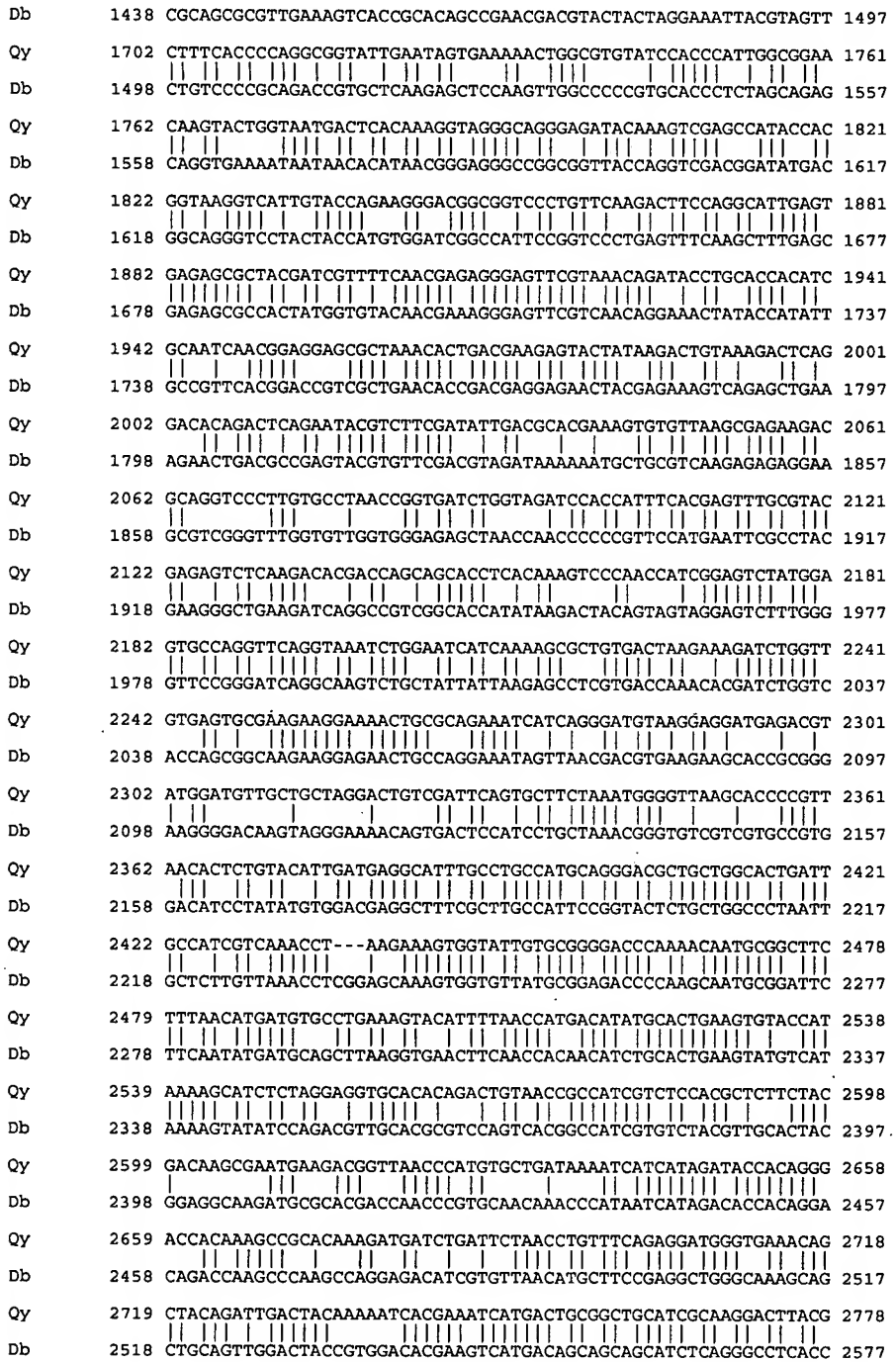
RESULT 8
US-09-190-246-1
; Sequence 1, Application US/09190246
; Publication No. US20030180257A1
; GENERAL INFORMATION:
; APPLICANT: Parrington, Mark
; APPLICANT: Li, Xiaomao
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: ALPHAVIRUS VECTORS FOR PARAMYXOVIRUS VACCINES
; FILE REFERENCE: Parrington et al.
; CURRENT APPLICATION NUMBER: US/09/190,246
; CURRENT FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 15538
; TYPE: DNA
; ORGANISM: Semliki Forest virus
US-09-190-246-1

Query Match 31.3%; Score 1564; DB 12; Length 15538;
Best Local Similarity 58.3%; Pred. No. 0;
Matches 2904; Conservative 0; Mismatches 2040; Indels 36; Gaps 8;


Db 2882 CCTATGCGCAGCGCAGAAGACCCCGAAAGGCTCGATAGCTACGCAAAGAAACTGGCAGCG 2941
Qy 325 AAGAAAAGTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACCTGCTGGAAGTC 384
Db 2942 GCCTCCGGGAAGGTGCTGGATAGAGAGATCGCAGGAAAAATCACCGACCTGCAGACCGTC 3001
Qy 385 ATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACGCCACGTGTAGG 444
Db 3002 ATGGCTACGCCAGACGCTGAATCTCCTACCTTTTGCCTGCATACAGACGTCACGTGTCGT 3061
Qy 445 TACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCACCGACATCAATC 504
Db 3062 ACGGCAGCCGAAGTGGCCGTATACCAGGACGTGTATGCTGTACATGCACCAACATCGCTG 3121
Qy 505 TACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTGACACGACCCCT 564
Db 3122 TACCATCAGGCGATGAAAGGTGTCAGAACGGCGTATTGGATTGGGTTTGACACCACCCCG 3181
Qy 565 TTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACTGGGCTGACGAG 624
Db 3182 TTTATGTTTGACGCGCTAGCAGGCGCGTATCCAACCTACGCCACAAACTGGGCCGACGAG 3241
Qy 625 AGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGGAGAGCAGGCTT 684
Db 3242 CAGGTGTTACAGGCCAGGAACATAGGACTGTGTGCAGCATCCTTGACTGAGGGAAGACTC 3301
Qy 685 GGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGATCATATTCTCG 744
Db 3302 GGCAAACTGTCCATTCTCCGCAAGAAGCAATTGAAACCTTGCGACACAGTCATGTTCTCG 3361
Qy 745 GTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGCATCTTCCAAAC 804
Db 3362 GTAGGATCTACATTGTACACTGAGAGCAGAAAGCTACTGAGGAGCTGGCACTTACCCTCC 3421
Qy 805 GTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTGTCAGCTGT 864
Db 3422 GTATTCCACCTGAAAGGTAAACAATCCTTTACCTGTAGGTGCGATACCATCGTATCATGT 3481
Qy 865 GAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAGTTGAGAAC 924
Db 3482 GAAGGGTACGTAGTTAAGAAAATCACTATGTGCCCCGGCCTGTACGGTAAAACGGTAGGG 3541
Qy 925 TTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATACGCTGCGC 984
Db 3542 TACGCCGTGACGTATCACGCGGAGGGATTCCTAGTGTGCAAGACCACAGACACTGTCAAA 3601
Qy 985 GGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCGATCAGATG 1044
Db 3602 GGAGAAAGAGTCTCATTCCCTGTATGCACCTACGTCCCCTCAACCATCTGTGATCAAATG 3661
Qy 1045 ACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTATTGGTTGGGCTC 1104
Db 3662 ACTGGCATACTAGCGACCGACGTCACACCGGAGGACGCACAGAAGTTGTTAGTGGGATTG 3721
Qy 1105 AACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAATGCAGAACTAT 1164
Db 3722 AATCAGAGGATAGTTGTGAACGGAAGAACACAGCGAAACACTAACACGATGAAGAACTAT 3781
Qy 1165 CTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATCGTGCCGACTTG 1224
Db 3782 CTGCTTCCGATTGTGGCCGTCGCATTTAGCAAGTGGGCGAGGGAATACAAGGCAGACCTT 3841
Qy 1225 GACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCTGCTGCTGGGCT 1284
Db 3842 GATGATGAAAAACCTCTGGGTGTCCGAGAGAGGTCACTTACTTGCTGCTGCTTGTGGGCA 3901
Qy 1285 TTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCTGGTACGCAAACAATTAAGAAA 1344
Db 3902 TTTAAAACGAGGAAGATGCACACCATGTACAAGAAACCAGACACCCAGACAATAGTGAAG 3961
Qy 1345 GTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCACGGGCTCGATATG 1404
Db 3962 GTGCCTTCAGAGTTTAACTCGTTCGTCATCCCGAGCCTATGGTCTACAGGCCTCGCAATC 4021
Qy 1405 GGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACTGTCAAACCCGCACCGGCTATT 1464
Db 4022 CCAGTCAGATCACGCATTAAGATGCTTTTGGCCAAGAAGACCAAGC---GAGAGTTAATA 4078
Qy 1465 ACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAGAAGTGGCTGCA 1524
Db 4079 CCTGTTCTCGACGCGTCGTCAGCCAGGGATGCTGAACAAGAGGAGAAGGAGAGGTTGGAG 4138
Qy 1525 GCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAAAAGAGACCGTA 1584
Db 4139 GCCGAGCTGACTAGAGAAGCCTTACCACCCCTCGTCCCCATCGCGCCGGCGGAGACGGGA 4198
Qy 1585 GAGGCAGAAGTAGACCTCATTATGCAAGAG------GCAGGAGCAGGTAGCGTGGAGACA 1638
Db 4199 GTCGTCGACGTCGACGTTGAAGAACTAGAGTATCACGCAGGTGCAGGGGTCGTGGAAACA 4258
Qy 1639 CCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAGATTGGGTCTTACGCT 1698
Db 4259 CCTCGCAGCGCGTTGAAAGTCACCGCACAGCCGAACGACGTACTACTAGGAAATTACGTA 4318
Qy 1699 ATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGTATCCACCCATTGGCG 1758
Db 4319 GTTCTGTCCCCGCAGACCGTGCTCAAGAGCTCCAAGTTGGCCCCCGTGCACCCTCTAGCA 4378
Qy 1759 GAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATACAAAGTCGAGCCATAC 1818
Db 4379 GAGCAGGTGAAAATAATAACACATAACGGGAGGGCCGGCGGTTACCAGGTCGACGGATAT 4438
Qy 1819 CACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAAGACTTCCAGGCATTG 1878
Db 4439 GACGGCAGGGTCCTACTACCATGTGGATCGGCCATTCCGGTCCCTGAGTTTCAAGCTTTG 4498
Qy 1879 AGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAACAGATACCTGCACCAC 1938
Db 4499 AGCGAGAGCGCCACTATGGTGTACAACGAAAGGGAGTTCGTCAACAGGAAACTATACCAT 4558
Qy 1939 ATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTATAAGACTGTAAAGACT 1998
Db 4559 ATTGCCGTTCACGGACCGTCGCTGAACACCGACGAGGAGAACTACGAGAAAGTCAGAGCT 4618
Qy 1999 CAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAGTGTGTTAAGCGAGAA 2058
Db 4619 GAAAGAACTGACGCCGAGTACGTGTTCGACGTAGATAAAAAATGCTGCGTCAAGAGAGAG 4678
Qy 2059 GACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCATTTCACGAGTTTGCG 2118
Db 4679 GAAGCGTCGGGTTTGGTGTTGGTGGGAGAGCTAACCAACCCCCCGTTCCATGAATTCGCC 4738
Qy 2119 TACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCAACCATCGGAGTCTAT 2178
Db 4739 TACGAAGGGCTGAAGATCAGGCCGTCGGCACCATATAAGACTACAGTAGTAGGAGTCTTT 4798
Qy 2179 GGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTGACTAAGAAAGATCTG 2238
Db 4799 GGGGTTCCGGGATCAGGCAAGTCTGCTATTATTAAGAGCCTCGTGACCAAACACGATCTG 4858
Qy 2239 GTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGATGTAAGGAGGATGAGA 2298
Db 4859 GTCACCAGCGGCAAGAAGGAGAACTGCCAGGAAATAGTTAACGACGTGAAGAAGCACCGC 4918
Qy 2299 CGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAATGGGGTTAAGCACCCC 2358
Db 4919 GGGAAGGGGACAAGTAGGGAAAACAGTGACTCCATCCTGCTAAACGGGTGTCGTCGTGCC 4978
Qy 2359 GTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGGACGCTGCTGGCACTG 2418
Db 4979 GTGGACATCCTATATGTGGACGAGGCTTTCGCTTGCCATTCCGGTACTCTGCTGGCCCTA 5038
Qy 2419 ATTGCCATCGTCAAACCT---AAGAAAGTGGTATTGTGCGGGGACCCAAAACAATGCGGC 2475
Db 5039 ATTGCTCTTGTTAAACCTCGGAGCAAAGTGGTGTTATGCGGAGACCCCAAGCAATGCGGA 5098


Qy 2476 TTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGCACTGAAGTGTAC 2535
Db 5099 TTCTTCAATATGATGCAGCTTAAGGTGAACTTCAACCACAACATCTGCACTGAAGTATGT 5158
Qy 2536 CATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATCGTCTCCACGCTCTTC 2595
Db 5159 CATAAAAGTATATCCAGACGTTGCACGCGTCCAGTCACGGCCATCGTGTCTACGTTGCAC 5218
Qy 2596 TACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATCATCATAGATACCACA 2655
Db 5219 TACGGAGGCAAGATGCGCACGACCAACCCGTGCAACAAACCCATAATCATAGACACCACA 5278
Qy 2656 GGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGAGGATGGGTGAAA 2715
Db 5279 GGACAGACCAAGCCCAAGCCAGGAGACATCGTGTTAACATGCTTCCGAGGCTGGGCAAAG 5338
Qy 2716 CAGCTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCTGCATCGCAAGGACTT 2775
Db 5339 CAGCTGCAGTTGGACTACCGTGGACACGAAGTCATGACAGCAGCAGCATCTCAGGGCCTC 5398
Qy 2776 ACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCACTCTACTCGCAG 2835
Db 5399 ACCCGCAAAGGGGTATACGCCGTAAGGCAGAAGGTGAATGAAAATCCCTTGTATGCCCCT 5458
Qy 2836 ACTTCTGAGCACGTGAACGTGTTACTTACACGCACAGAAAAACGCATTGTCTGGAAGACG 2895
Db 5459 GCGTCGGAGCACGTGAATGTACTGCTGACGCGCACTGAGGATAGGCTGGTGTGGAAAACG 5518
Qy 2896 CTAGCTGGTGATCCCTGGATAAAGACACTTACAGCTAAATATCCCGGGGATTTCACGGCT 2955
Db 5519 CTGGCCGGCGATCCCTGGATTAAGGTCCTATCAAACATTCCACAGGGTAACTTTACGGCC 5578
Qy 2956 TCATTGGACGACTGGCAGCGCGAACACGACGCCATTATGGCACGCGTTCTTGATAAGCCG 3015
Db 5579 ACATTGGAAGAATGGCAAGAAGAACACGACAAAATAATGAAGGTGATTGAAGGACCGGCT 5638
Qy 3016 CAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAGGCTTTAGAGCCA 3075
Db 5639 GCGCCTGTGGACGCGTTCCAGAACAAAGCGAACGTGTGTTGGGCGAAAAGCCTGGTGCCT 5698
Qy 3076 GTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAGAC---GTTGCACCCA 3132
Db 5699 GTCCTGGACACTGCCGGAATCAGATTGACAGCAGAGGAGTGGAGCACCATAATTACAGCA 5758
Qy 3133 TTCAAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTTTGCACCAGGTTC 3192
Db 5759 TTTAAGGAGGACAGAGCTTACTCTCCAGTGGTGGCCTTGAATGAAATTTGCACCAAGTAC 5818
Qy 3193 TTTGGAGTAGACCTGGACAGTGGGTTATTTTCCGCTCCTACCGTCGCACTTACTTACAGG 3252
Db 5819 TATGGAGTTGACCTGGACAGTGGCCTGTTTTCTGCCCCGAAGGTGTCCCTGTATTACGAG 5878
Qy 3253 GATCAGCACTGGGATAACTCGCCAGGGAAGAACATGTATGGGCTTAATAGAGAGGTAGCA 3312
Db 5879 AACAACCACTGGGATAACAGACCTGGTGGAAGGATGTATGGATTCAATGCCGCAACAGCT 5938
Qy 3313 AAGGAGTTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACAGGCAGGGTAGCT 3372
Db 5939 GCCAGGCTGGAAGCTAGACATACCTTCCTGAAGGGGCAGTGGCATACGGGCAAGCAGGCA 5998
Qy 3373 GATATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTGGTTCCATTAAAT 3432
Db 5999 GTTATCGCAGAAAGAAAAATCCAACCGCTTTCTGTGCTGGACAATGTAATTCCTATCAAC 6058
Qy 3433 CGCCGGTTGCCCCACTCGTTGATCGTTGACCACAAAGGACAGGGTACAACTGATCACAGC 3492
Db 6059 CGCAGGCTGCCGCACGCCCTGGTGGCTGAGTACAAGACGGTTAAAGGCAGTAGGGTTGAG 6118
Qy 3493 GGATTCCTATCTAAGATGAAGGGCAAATCTGTGTTGGTGATCGGCGATCCTATCAGCATT 3552
Db 6119 TGGCTGGTCAATAAAGTAAGAGGGTACCACGTCCTGCTGGTGAGTGAGTACAACCTGGCT 6178
Qy 3553 CCAGGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATAC---------CATCAGG 3603
Db 6179 TTGCCTCGACGCAGGGTCACTTGGTTGTCACCGCTGAATGTCACAGGCGCCGATAGGTGC 6238
Qy 3604 TGTGATCTCGATTTGGGAATACCTAGCCATGTCGGTAAATATGACATTATCTTTGTCAAT 3663
Db 6239 TACGACCTAAGTTTAGGACTGCCGGCTGACGCCGGCAGGTTCGACTTGGTCTTTGTGAAC 6298
Qy 3664 GTTAGGACCCCGTACAGGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCACCAC 3723
Db 6299 ATTCACACGGAATTCAGAATCCACCACTACCAGCAGTGTGTCGACCACGCCATGAAGCTG 6358
Qy 3724 AGCATGCTAACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCTATA 3783
Db 6359 CAGATGCTTGGGGGAGATGCGCTACGACTGCTAAAACCCGGCGGCATCT---TGATGAGA 6415
Qy 3784 GGGTATGGGCTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCACGCTCATTT 3843
Db 6416 GCTTACGGATACGCCGATAAAATCAGCGAAGCCGTTGTTTCCTCCTTAAGCAGAAAGTTC 6475
Qy 3844 AGGTTTACCCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTCGTG 3903
Db 6476 TCGTCTGCAAGAGTGTTGCGCCCGGATTGTGTCACCAGCAATACAGAAGTGTTCTTGCTG 6535
Qy 3904 TTCTTCGGCAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTGCTT 3963
Db 6536 TTCTCCAACTTTGACAACGGAAAGAGACCCTCTACGCTACACCAGATGAATACCAAGCTG 6595
Qy 3964 GACAACATCTATCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGTACAGAGTG 4023
Db 6596 AGTGCCGTGTATGCCGGAGAAGCCATGCACACGGCCGGGTGTGCACCATCCTACAGAGTT 6655
Qy 4024 ATCAGAGGTGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGCAAA 4083
Db 6656 AAGAGAGCAGACATAGCCACGTGCACAGAAGCGGCTGTGGTTAACGCAGCTAACGCCCGT 6715
Qy 4084 GGTCAACCAGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTTGAT 4143
Db 6716 GGAACTGTAGGGGATGGCGTATGCAGGGCCGTGGCGAAGAAATGGCCGTCAGCCTTTAAG 6775
Qy 4144 AGACAGCCAATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACCG---CTCATCATA 4200
Db 6776 GGAGCAGCAACACCAGTGGGCACAATTAAAACAGTCATGTGCGGCTCGTACCCCGTCATC 6835
Qy 4201 CATGCTGTAGGACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTAAGCTCGCA 4260
Db 6836 CACGCTGTAGCGCCTAATTTCTCTGCCACGACTGAAGCGGAAGGGGACCGCGAATTGGCC 6895
Qy 4261 GCTGCCTACATGAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTA 4320
Db 6896 GCTGTCTACCGGGCAGTGGCCGCCGAAGTAAACAGACTGTCACTGAGCAGCGTAGCCATC 6955
Qy 4321 CCGCTACTGTCAACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCAT 4380
Db 6956 CCGCTGCTGTCCACAGGAGTGTTCAGCGGCGGAAGAGATAGGCTGCAGCAATCCCTCAAC 7015
Qy 4381 CACCTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAA 4440
Db 7016 CATCTATTCACAGCAATGGACGCCACGGACGCTGACGTGACCATCTACTGCAGAGACAAA 7075
Qy 4441 CAATGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATTCTGGAT 4500
Db 7076 AGTTGGGAGAAGAAAATCCAGGAAGCCATTGACATGAGGACGGCTGTGGAGTTGCTCAAT 7135
Qy 4501 GATGACAAGCCAGTAGACATTGACTTGGTCAGGGTCCACCCAAACAGCTCTTTGGCAGGC 4560
Db 7136 GATGACGTGGAGCTGACCACAGACTTGGTGAGAGTGCACCCGGACAGCAGCCTGGTGGGT 7195
Qy 4561 AGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTGGAAGGTACACGATTC 4620
Db 7196 CGTAAGGGCTACAGTACCACTGACGGGTCGCTGTACTCGTACTTTGAAGGTACGAAATTC 7255
Qy 4621 CATCAGACCGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCCAACAAATCTGAGGCT 4680

```
Db       7256 AACCAGGCTGCTATTGATATGGCAGAGATACTGACGTTGTGGCCCAGACTGCAAGAGGCA 7315

Qy       4681 AATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGC 4740

Db       7316 AACGAACAGATATGCCTATACGCGCTGGGCGAAACAATGGACAACATCAGATCCAAATGT 7375

Qy       4741 CCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTAC 4800

Db       7376 CCGGTGAACGATTCCGATTCATCAACACCTCCCAGGACAGTGCCCTGCCTGTGCCGCTAC 7435

Qy       4801 GCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTA 4860

Db       7436 GCAATGACAGCAGAACGGATCGCCCGCCTTAGGTCACACCAAGTTAAAAGCATGGTGGTT 7495

Qy       4861 TGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGC 4920

Db       7496 TGCTCATCTTTTCCCCTCCCGAAATACCATGTAGATGGGGTGCAGAAGGTAAAGTGCGAG 7555

Qy       4921 AAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGCGGAA 4980

Db       7556 AAGGTTCTCCTGTTCGACCCGACGGTACCTTCAGTGGTTAGTCCGCGGAAGTATGCCGCA 7615


RESULT 9
US-09-190-246-4
; Sequence 4, Application US/09190246
; Publication No. US20030180257A1
; GENERAL INFORMATION:
;  APPLICANT: Parrington, Mark
;  APPLICANT: Li, Xiaomao
;  APPLICANT: Klein, Michel
;  TITLE OF INVENTION: ALPHAVIRUS VECTORS FOR PARAMYXOVIRUS VACCINES
;  FILE REFERENCE: Parrington et al.
;  CURRENT APPLICATION NUMBER: US/09/190,246
;  CURRENT FILING DATE: 1998-11-13
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 8100
;   TYPE: DNA
;   ORGANISM: Semliki Forest virus
US-09-190-246-4

  Query Match             29.5%;  Score 1473.8;  DB 12;  Length 8100;
  Best Local Similarity   57.9%;  Pred. No. 0;
  Matches 2779;  Conservative    0;  Mismatches 1982;  Indels   36;  Gaps    8;

Qy        214 ATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACCGCTATCATTGTATCTGCCCT 273

Db          1 ATCGGCAGTGCGCCTTCCAGGAGAATGATGTCTACGCACAAATACCACTGCGTATGCCCT 60

Qy        274 ATGATAAGCGCTGAAGACCCGGACAGACTACAACGGTATGCAGAAAGACTTAAGAAAA-- 331

Db         61 ATGCGCAGCGCAGAAGACCCCGAAAGGCTCGATAGCTACGCAAAGAAACTGGCAGCGGCC 120

Qy        332 ----GTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACCTGCTGGAAGTCATG 387

Db        121 TCCGGGAAGGTGCTGGATAGAGAGATCGCAGGAAAAATCACCGACCTGCAGACCGTCATG 180

Qy        388 TCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACGCCACGTGTAGGTAC 447

Db        181 GCTACGCCAGACGCTGAATCTCCTACCTTTTGCCTGCATACAGACGTCACGTGTCGTACG 240

Qy        448 TTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCACCGACATCAATCTAC 507

Db        241 GCAGCCGAAGTGGCCGTATACCAGGACGTGTATGCTGTACATGCACCAACATCGCTGTAC 300

Qy        508 CACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTGACACGACCCCTTTT 567

Db        301 CATCAGGCGATGAAAGGTGTCAGAACGGCGTATTGGATTGGGTTTGACACCACCCCGTTT 360

Qy        568 ATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACTGGGCTGACGAGAGA 627

Db        361 ATGTTTGACGCGCTAGCAGGCGCGTATCCAACCTACGCCACAAACTGGGCCGACGAGCAG 420

Qy        628 GTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGGAGAGCAGGCTTGGA 687

Db        421 GTGTTACAGGCCAGGAACATAGGACTGTGTGCAGCATCCTTGACTGAGGGAAGACTCGGC 480

Qy        688 AAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGATCATATTCTCGGTT 747

Db        481 AAACTGTCCATTCTCCGCAAGAAGCAATTGAAACCTTGCGACACAGTCATGTTCTCGGTA 540

Qy        748 GGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGCATCTTCCAAACGTG 807

Db        541 GGATCTACATTGTACACTGAGAGCAGAAAGCTACTGAGGAGCTGGCACTTACCCTCCGTA 600

Qy        808 TTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTGTCAGCTGTGAA 867

Db        601 TTCCACCTGAAAGGTAAACAATCCTTTACCTGTAGGTGCGATACCATCGTATCATGTGAA 660

Qy        868 GGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAGTTGAGAACTTG 927

Db        661 GGGTACGTAGTTAAGAAAATCACTATGTGCCCCGGCCTGTACGGTAAAACGGTAGGGTAC 720

Qy        928 GCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATACGCTGCGCGGC 987

Db        721 GCCGTGACGTATCACGCGGAGGGATTCCTAGTGTGCAAGACCACAGACACTGTCAAAGGA 780

Qy        988 GAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCGATCAGATGACA 1047

Db        781 GAAAGAGTCTCATTCCCTGTATGCACCTACGTCCCCTCAACCATCTGTGATCAAATGACT 840

Qy       1048 GGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTATTGGTTGGGCTCAAC 1107

Db        841 GGCATACTAGCGACCGACGTCACACCGGAGGACGCACAGAAGTTGTTAGTGGGATTGAAT 900

Qy       1108 CAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAATGCAGAACTATCTA 1167

Db        901 CAGAGGATAGTTGTGAACGGAAGAACACAGCGAAACACTAACACGATGAAGAACTATCTG 960

Qy       1168 TTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATCGTGCCGACTTGGAC 1227

Db        961 CTTCCGATTGTGGCCGTCGCATTTAGCAAGTGGGCGAGGGAATACAAGGCAGACCTTGAT 1020

Qy       1228 GACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCTGCTGCTGGGCTTTC 1287

Db       1021 GATGAAAAACCTCTGGGTGTCCGAGAGAGGTCACTTACTTGCTGCTGCTTGTGGGCATTT 1080

Qy       1288 AAGACCCAGAAAATCACATCCATCTACAAGAAGCCTGGTACGCAAACAATTAAGAAAGTA 1347

Db       1081 AAAACGAGGAAGATGCACACCATGTACAAGAAACCAGACACCCAGACAATAGTGAAGGTG 1140

Qy       1348 CCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCACGGGCTCGATATGGGC 1407

Db       1141 CCTTCAGAGTTTAACTCGTTCGTCATCCCGAGCCTATGGTCTACAGGCCTCGCAATCCCA 1200

Qy       1408 TTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACTGTCAAACCCGCACCGGCTATTACA 1467

Db       1201 GTCAGATCACGCATTAAGATGCTTTTGGCCAAGAAGACCAAGCGAGA---GTTAATACCT 1257

Qy       1468 ATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAGAAGTGGCTGCAGCG 1527

Db       1258 GTTCTCGACGCGTCGTCAGCCAGGGATGCTGAACAAGAGGAGAAGGAGAGGTTGGAGGCC 1317

Qy       1528 GAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAAAAGAGACCGTAGAG 1587

Db       1318 GAGCTGACTAGAGAAGCCTTACCACCCCTCGTCCCCATCGCGCCGGCGGAGACGGGAGTC 1377

Qy       1588 GCAGAAGTAGACCTCATTATGCAAGAG------GCAGGAGCAGGTAGCGTGGAGACACCA 1641

Db       1378 GTCGACGTCGACGTTGAAGAACTAGAGTATCACGCAGGTGCAGGGGTCGTGGAAACACCT 1437

Qy       1642 CGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAGATTGGGTCTTACGCTATA 1701
```


Db 1438 CGCAGCGCGTTGAAAGTCACCGCACAGCCGAACGACGTACTACTAGGAAATTACGTAGTT 1497
Qy 1702 CTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGTATCCACCCATTGGCGGAA 1761
Db 1498 CTGTCCCCGCAGACCGTGCTCAAGAGCTCCAAGTTGGCCCCCGTGCACCCTCTAGCAGAG 1557
Qy 1762 CAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATACAAAGTCGAGCCATACCAC 1821
Db 1558 CAGGTGAAAATAATAACACATAACGGGAGGGCCGGCGGTTACCAGGTCGACGGATATGAC 1617
Qy 1822 GGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAAGACTTCCAGGCATTGAGT 1881
Db 1618 GGCAGGGTCCTACTACCATGTGGATCGGCCATTCCGGTCCCTGAGTTTCAAGCTTTGAGC 1677
Qy 1882 GAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAACAGATACCTGCACCACATC 1941
Db 1678 GAGAGCGCCACTATGGTGTACAACGAAAGGGAGTTCGTCAACAGGAAACTATACCATATT 1737
Qy 1942 GCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTATAAGACTGTAAAGACTCAG 2001
Db 1738 GCCGTTCACGGACCGTCGCTGAACACCGACGAGGAGAACTACGAGAAAGTCAGAGCTGAA 1797
Qy 2002 GACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAGTGTGTTAAGCGAGAAGAC 2061
Db 1798 AGAACTGACGCCGAGTACGTGTTCGACGTAGATAAAAAATGCTGCGTCAAGAGAGAGGAA 1857
Qy 2062 GCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCATTTCACGAGTTTGCGTAC 2121
Db 1858 GCGTCGGGTTTGGTGTTGGTGGGAGAGCTAACCAACCCCCCGTTCCATGAATTCGCCTAC 1917
Qy 2122 GAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCAACCATCGGAGTCTATGGA 2181
Db 1918 GAAGGGCTGAAGATCAGGCCGTCGGCACCATATAAGACTACAGTAGTAGGAGTCTTTGGG 1977
Qy 2182 GTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTGACTAAGAAAGATCTGGTT 2241
Db 1978 GTTCCGGGATCAGGCAAGTCTGCTATTATTAAGAGCCTCGTGACCAAACACGATCTGGTC 2037
Qy 2242 GTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGATGTAAGGAGGATGAGACGT 2301
Db 2038 ACCAGCGGCAAGAAGGAGAACTGCCAGGAAATAGTTAACGACGTGAAGAAGCACCGCGGG 2097
Qy 2302 ATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAATGGGGTTAAGCACCCCGTT 2361
Db 2098 AAGGGGACAAGTAGGGAAAACAGTGACTCCATCCTGCTAAACGGGTGTCGTCGTGCCGTG 2157
Qy 2362 AACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGGACGCTGCTGGCACTGATT 2421
Db 2158 GACATCCTATATGTGGACGAGGCTTTCGCTTGCCATTCCGGTACTCTGCTGGCCCTAATT 2217
Qy 2422 GCCATCGTCAAACCT---AAGAAAGTGGTATTGTGCGGGGACCCAAAACAATGCGGCTTC 2478
Db 2218 GCTCTTGTTAAACCTCGGAGCAAAGTGGTGTTATGCGGAGACCCCAAGCAATGCGGATTC 2277
Qy 2479 TTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGCACTGAAGTGTACCAT 2538
Db 2278 TTCAATATGATGCAGCTTAAGGTGAACTTCAACCACAACATCTGCACTGAAGTATGTCAT 2337
Qy 2539 AAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATCGTCTCCACGCTCTTCTAC 2598
Db 2338 AAAAGTATATCCAGACGTTGCACGCGTCCAGTCACGGCCATCGTGTCTACGTTGCACTAC 2397
Qy 2599 GACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATCATCATAGATACCACAGGG 2658
Db 2398 GGAGGCAAGATGCGCACGACCAACCCGTGCAACAAACCCATAATCATAGACACCACAGGA 2457
Qy 2659 ACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGAGGATGGGTGAAACAG 2718
Db 2458 CAGACCAAGCCCAAGCCAGGAGACATCGTGTTAACATGCTTCCGAGGCTGGGCAAAGCAG 2517
Qy 2719 CTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCTGCATCGCAAGGACTTACG 2778
Db 2518 CTGCAGTTGGACTACCGTGGACACGAAGTCATGACAGCAGCAGCATCTCAGGGCCTCACC 2577
Qy 2779 CGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCACTCTACTCGCAGACT 2838
Db 2578 CGCAAAGGGGTATACGCCGTAAGGCAGAAGGTGAATGAAAATCCCTTGTATGCCCCTGCG 2637
Qy 2839 TCTGAGCACGTGAACGTGTTACTTACACGCACAGAAAAACGCATTGTCTGGAAGACGCTA 2898
Db 2638 TCGGAGCACGTGAATGTACTGCTGACGCGCACTGAGGATAGGCTGGTGTGGAAAACGCTG 2697
Qy 2899 GCTGGTGATCCCTGGATAAAGACACTTACAGCTAAATATCCCGGGGATTTCACGGCTTCA 2958
Db 2698 GCCGGCGATCCCTGGATTAAGGTCCTATCAAACATTCCACAGGGTAACTTTACGGCCACA 2757
Qy 2959 TTGGACGACTGGCAGCGCGAACACGACGCCATTATGGCACGCGTTCTTGATAAGCCGCAG 3018
Db 2758 TTGGAAGAATGGCAAGAAGAACACGACAAAATAATGAAGGTGATTGAAGGACCGGCTGCG 2817
Qy 3019 ACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAGGCTTTAGAGCCAGTC 3078
Db 2818 CCTGTGGACGCGTTCCAGAACAAAGCGAACGTGTGTTGGGCGAAAAGCCTGGTGCCTGTC 2877
Qy 3079 TTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAGAC---GTTGCACCCATTC 3135
Db 2878 CTGGACACTGCCGGAATCAGATTGACAGCAGAGGAGTGGAGCACCATAATTACAGCATTT 2937
Qy 3136 AAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTTTGCACCAGGTTCTTT 3195
Db 2938 AAGGAGGACAGAGCTTACTCTCCAGTGGTGGCCTTGAATGAAATTTGCACCAAGTACTAT 2997
Qy 3196 GGAGTAGACCTGGACAGTGGGTTATTTTCCGCTCCTACCGTCGCACTTACTTACAGGGAT 3255
Db 2998 GGAGTTGACCTGGACAGTGGCCTGTTTTCTGCCCCGAAGGTGTCCCTGTATTACGAGAAC 3057
Qy 3256 CAGCACTGGGATAACTCGCCAGGGAAGAACATGTATGGGCTTAATAGAGAGGTAGCAAAG 3315
Db 3058 AACCACTGGGATAACAGACCTGGTGGAAGGATGTATGGATTCAATGCCGCAACAGCTGCC 3117
Qy 3316 GAGTTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACAGGCAGGGTAGCTGAT 3375
Db 3118 AGGCTGGAAGCTAGACATACCTTCCTGAAGGGGCAGTGGCATACGGGCAAGCAGGCAGTT 3177
Qy 3376 ATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTGGTTCCATTAAATCGC 3435
Db 3178 ATCGCAGAAAGAAAAATCCAACCGCTTTCTGTGCTGGACAATGTAATTCCTATCAACCGC 3237
Qy 3436 CGGTTGCCCCACTCGTTGATCGTTGACCACAAAGGACAGGGTACAACTGATCACAGCGGA 3495
Db 3238 AGGCTGCCGCACGCCCTGGTGGCTGAGTACAAGACGGTTAAAGGCAGTAGGGTTGAGTGG 3297
Qy 3496 TTCCTATCTAAGATGAAGGGCAAATCTGTGTTGGTGATCGGCGATCCTATCAGCATTCCA 3555
Db 3298 CTGGTCAATAAAGTAAGAGGGTACCACGTCCTGCTGGTGAGTGAGTACAACCTGGCTTTG 3357
Qy 3556 GGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATAC---------CATCAGGTGT 3606
Db 3358 CCTCGACGCAGGGTCACTTGGTTGTCACCGCTGAATGTCACAGGCGCCGATAGGTGCTAC 3417
Qy 3607 GATCTCGATTTGGGAATACCTAGCCATGTCGGTAAATATGACATTATCTTTGTCAATGTT 3666
Db 3418 GACCTAAGTTTAGGACTGCCGGCTGACGCCGGCAGGTTCGACTTGGTCTTTGTGAACATT 3477
Qy 3667 AGGACCCCGTACAGGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCACCACAGC 3726
Db 3478 CACACGGAATTCAGAATCCACCACTACCAGCAGTGTGTCGACCACGCCATGAAGCTGCAG 3537
Qy 3727 ATGCTAACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCTATAGGG 3786
Db 3538 ATGCTTGGGGGAGATGCGCTACGACTGCTAAAACCCGGCGGCATCT---TGATGAGAGCT 3594
Qy 3787 TATGGGCTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCACGCTCATTTAGG 3846
Db 3595 TACGGATACGCCGATAAAATCAGCGAAGCCGTTGTTTCCTCCTTAAGCAGAAAGTTCTCG 3654

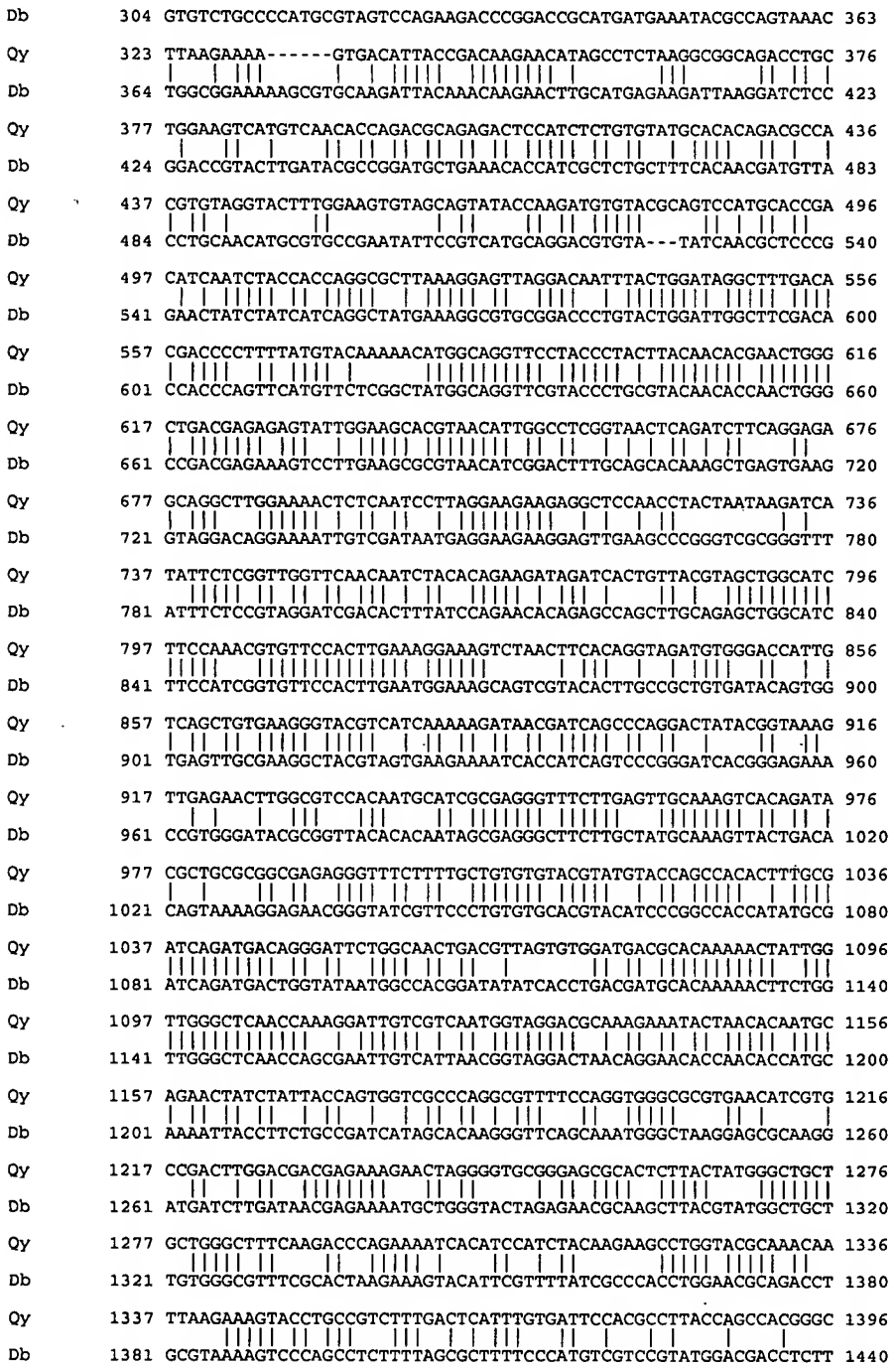
```
Qy       3847 TTTACCCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTCGTGTTC 3906
Db       3655 TCTGCAAGAGTGTTGCGCCCGGATTGTGTCACCAGCAATACAGAAGTGTTCTTGCTGTTC 3714

Qy       3907 TTCGGCAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTGCTTGAC 3966
Db       3715 TCCAACTTTGACAACGGAAAGAGACCCTCTACGCTACACCAGATGAATACCAAGCTGAGT 3774

Qy       3967 AACATCTATCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGTACAGAGTGATC 4026
Db       3775 GCCGTGTATGCCGGAGAAGCCATGCACACGGCCGGGTGTGCACCATCCTACAGAGTTAAG 3834

Qy       4027 AGAGGTGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGCAAAGGT 4086
Db       3835 AGAGCAGACATAGCCACGTGCACAGAAGCGGCTGTGGTTAACGCAGCTAACGCCCGTGGA 3894

Qy       4087 CAACCAGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTTGATAGA 4146
Db       3895 ACTGTAGGGGATGGCGTATGCAGGGCCGTGGCGAAGAAATGGCCGTCAGCCTTTAAGGGA 3954

Qy       4147 CAGCCAATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACCG---CTCATCATACAT 4203
Db       3955 GCAGCAACACCAGTGGGCACAATTAAAACAGTCATGTGCGGCTCGTACCCCGTCATCCAC 4014

Qy       4204 GCTGTAGGACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTAAGCTCGCAGCT 4263
Db       4015 GCTGTAGCGCCTAATTTCTCTGCCACGACTGAAGCGGAAGGGGACCGCGAATTGGCCGCT 4074

Qy       4264 GCCTACATGAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTACCG 4323
Db       4075 GTCTACCGGGCAGTGGCCGCCGAAGTAAACAGACTGTCACTGAGCAGCGTAGCCATCCCG 4134

Qy       4324 CTACTGTCAACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCATCAC 4383
Db       4135 CTGCTGTCCACAGGAGTGTTCAGCGGCGGAAGAGATAGGCTGCAGCAATCCCTCAACCAT 4194

Qy       4384 CTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAACAA 4443
Db       4195 CTATTCACAGCAATGGACGCCACGGACGCTGACGTGACCATCTACTGCAGAGACAAAAGT 4254

Qy       4444 TGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATTCTGGATGAT 4503
Db       4255 TGGGAGAAGAAAATCCAGGAAGCCATTGACATGAGGACGGCTGTGGAGTTGCTCAATGAT 4314

Qy       4504 GACAAGCCAGTAGACATTGACTTGGTCAGGGTCCACCCAAACAGCTCTTTGGCAGGCAGA 4563
Db       4315 GACGTGGAGCTGACCACAGACTTGGTGAGAGTGCACCCGGACAGCAGCCTGGTGGGTCGT 4374

Qy       4564 CCAGGTTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTGGAAGGTACACGATTCCAT 4623
Db       4375 AAGGGCTACAGTACCACTGACGGGTCGCTGTACTCGTACTTTGAAGGTACGAAATTCAAC 4434

Qy       4624 CAGACCGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCCAACAAATCTGAGGCTAAT 4683
Db       4435 CAGGCTGCTATTGATATGGCAGAGATACTGACGTTGTGGCCCAGACTGCAAGAGGCAAAC 4494

Qy       4684 GAGCAGATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGCCCA 4743
Db       4495 GAACAGATATGCCTATACGCGCTGGGCGAAACAATGGACAACATCAGATCCAAATGTCCG 4554

Qy       4744 GTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTACGCT 4803
Db       4555 GTGAACGATTCCGATTCATCAACACCTCCCAGGACAGTGCCCTGCCTGTGCCGCTACGCA 4614

Qy       4804 ATGACGGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTATGC 4863
Db       4615 ATGACAGCAGAACGGATCGCCCGCCTTAGGTCACACCAAGTTAAAAGCATGGTGGTTTGC 4674

Qy       4864 TCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGCAAA 4923
Db       4675 TCATCTTTTCCCCTCCCGAAATACCATGTAGATGGGGTGCAGAAGGTAAAGTGCGAGAAG 4734

Qy       4924 CCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGCGGAA 4980
```

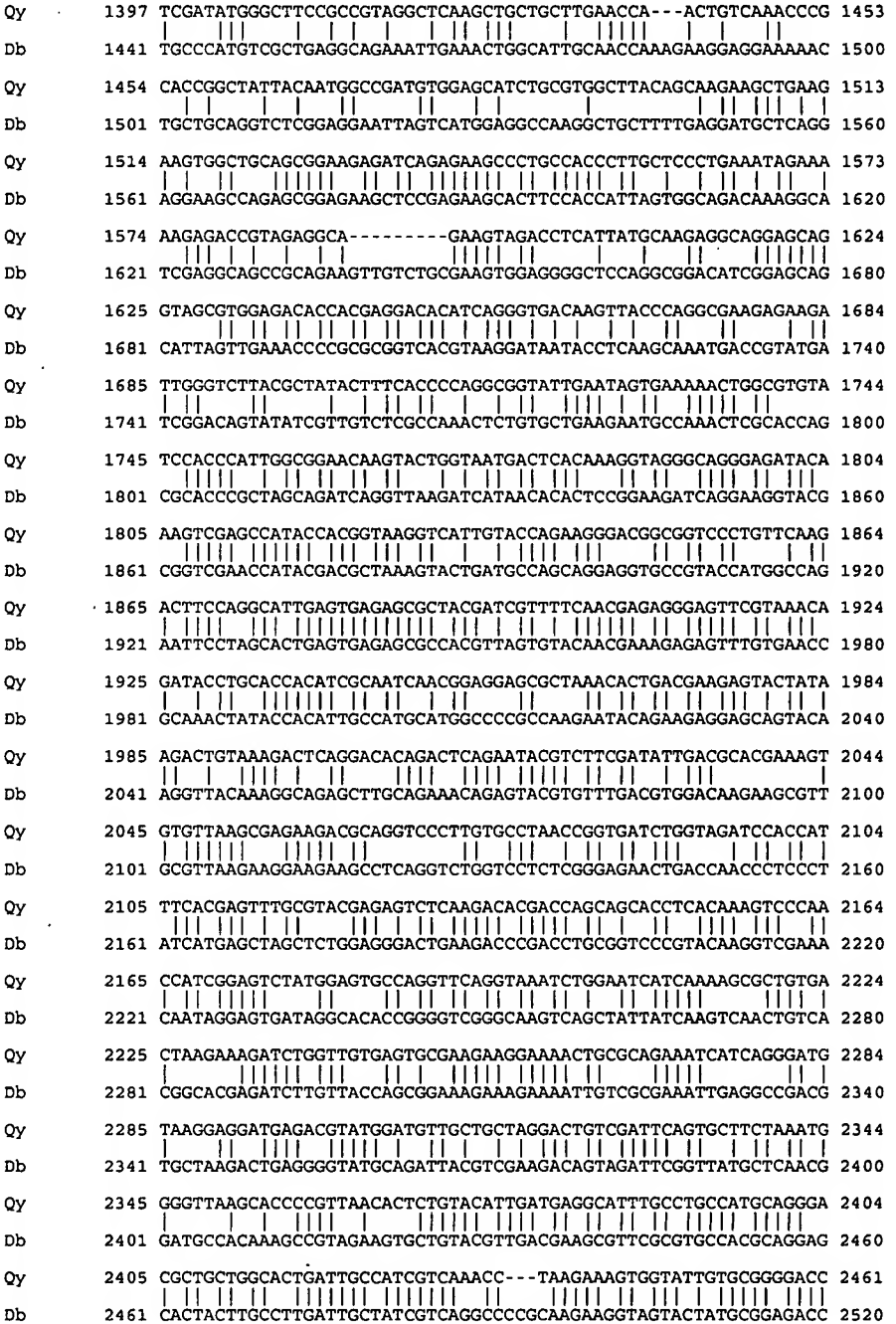
```
Db       4735 GTTCTCCTGTTCGACCCGACGGTACCTTCAGTGGTTAGTCCGCGGAAGTATGCCGCA 4791
```

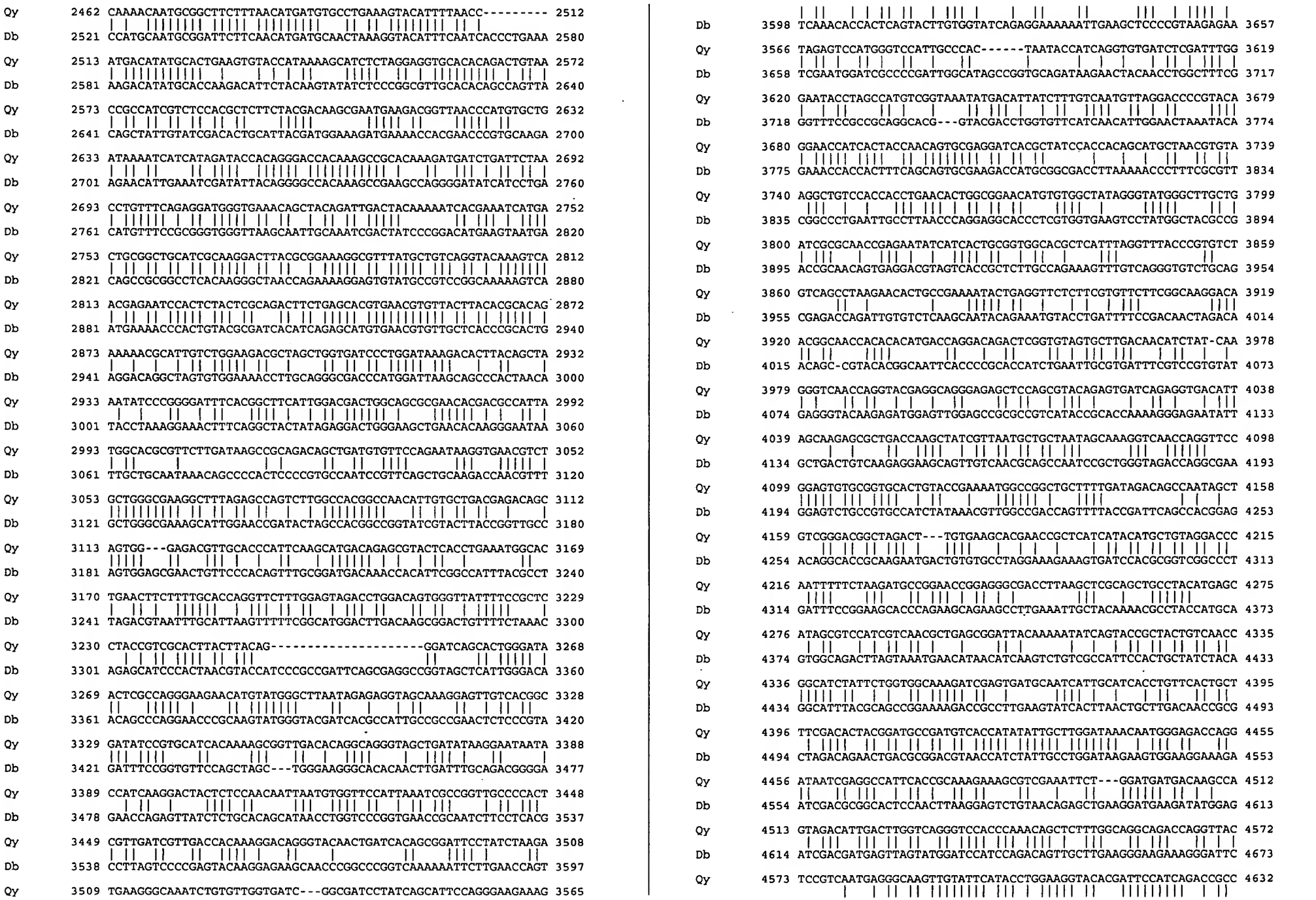
RESULT 10
US-09-507-362-102
; Sequence 102, Application US/09507362
; Publication No. US20030096397A1
; GENERAL INFORMATION:
; APPLICANT: Dubensky Jr., Thomas W.
; Polo, John M.
; Belli, Barbara A.
; Schlesinger, Sondra
; Dryga, Sergey A.
; Frolov, Ilya
; TITLE OF INVENTION: RECOMBINANT ALPHAVIRUS-BASED VECTORS
; WITH REDUCED INHIBITION OF CELLULAR MACRO-MOLECULAR
; SYNTHESIS
; NUMBER OF SEQUENCES: 125
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed Intellectual Property Law Group PLLC
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/507,362
; FILING DATE: 18-Feb-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 930049.457D6 /1196.011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8000 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-09-507-362-102

Query Match 26.0%; Score 1300.4; DB 11; Length 8000;
Best Local Similarity 56.0%; Pred. No. 0;
Matches 2819; Conservative 0; Mismatches 2131; Indels 80; Gaps 16;

```
Qy         23 ATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGTCGTTACAGC 82
Db         64 AGAAGCCAGTAGTAAACGTAGACGTAGACCCCCAGAGTCCGTTTGTCGTGCAACTGCAAA 123

Qy         83 GGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACCATGCCAATG 142
Db        124 AAAGCTTCCCGCAATTTGAGGTAGTAGCACAGCAGGTCACTCCAAATGACCATGCTAATG 183

Qy        143 CCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACCGGGACCAAG 202
Db        184 CCAGAGCATTTTCGCATCTGGCCAGTAAACTAATCGAGCTGGAGGTTCCTACCACAGCGA 243

Qy        203 TTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACCGCTATCATT 262
Db        244 CGATCTTGGACATAGGCAGCGCACCGGCTCGTAGAATGTTTTCCGAGCACCAGTATCATT 303

Qy        263 GTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGGTATGCAGAAAGAC 322
```





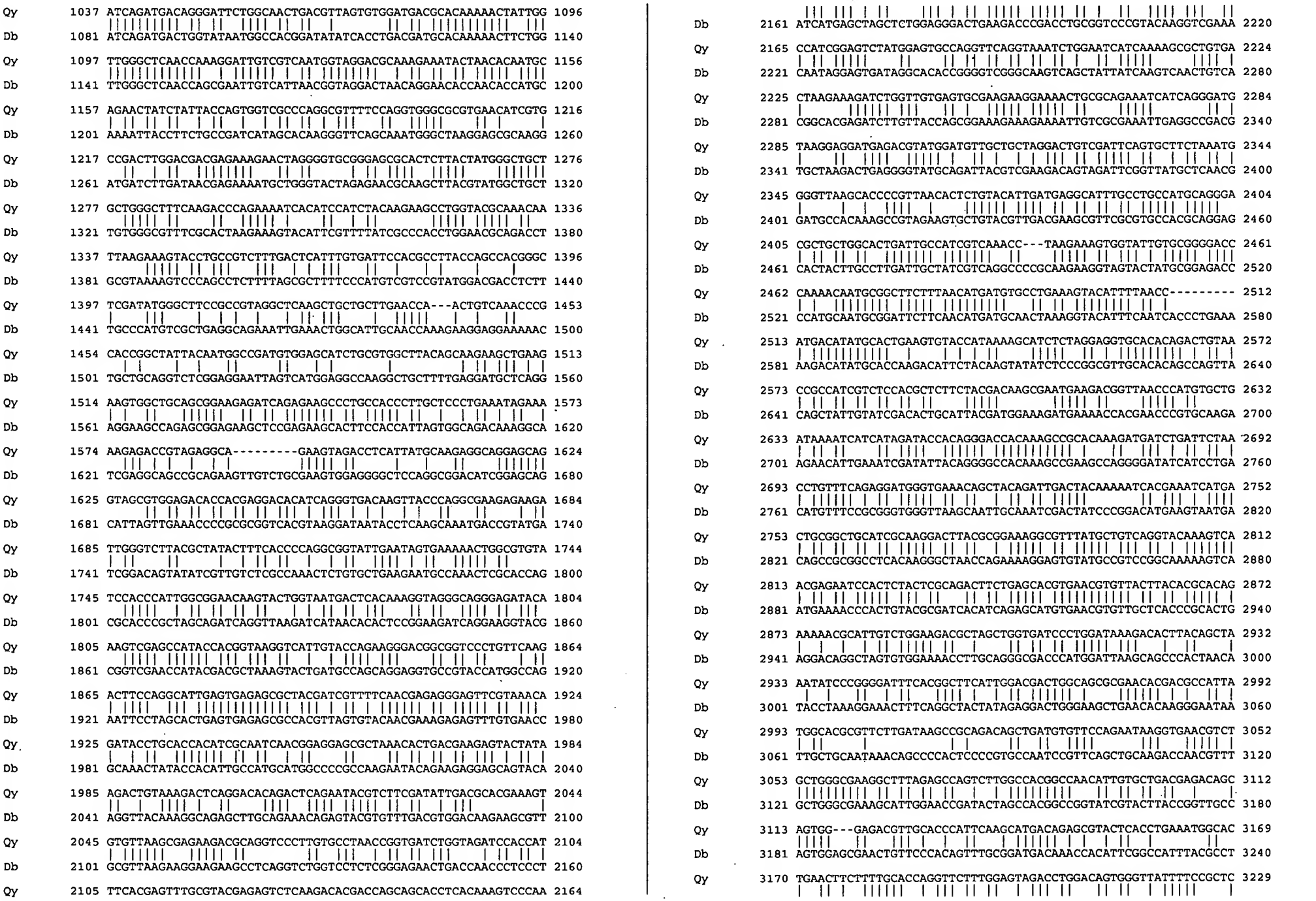


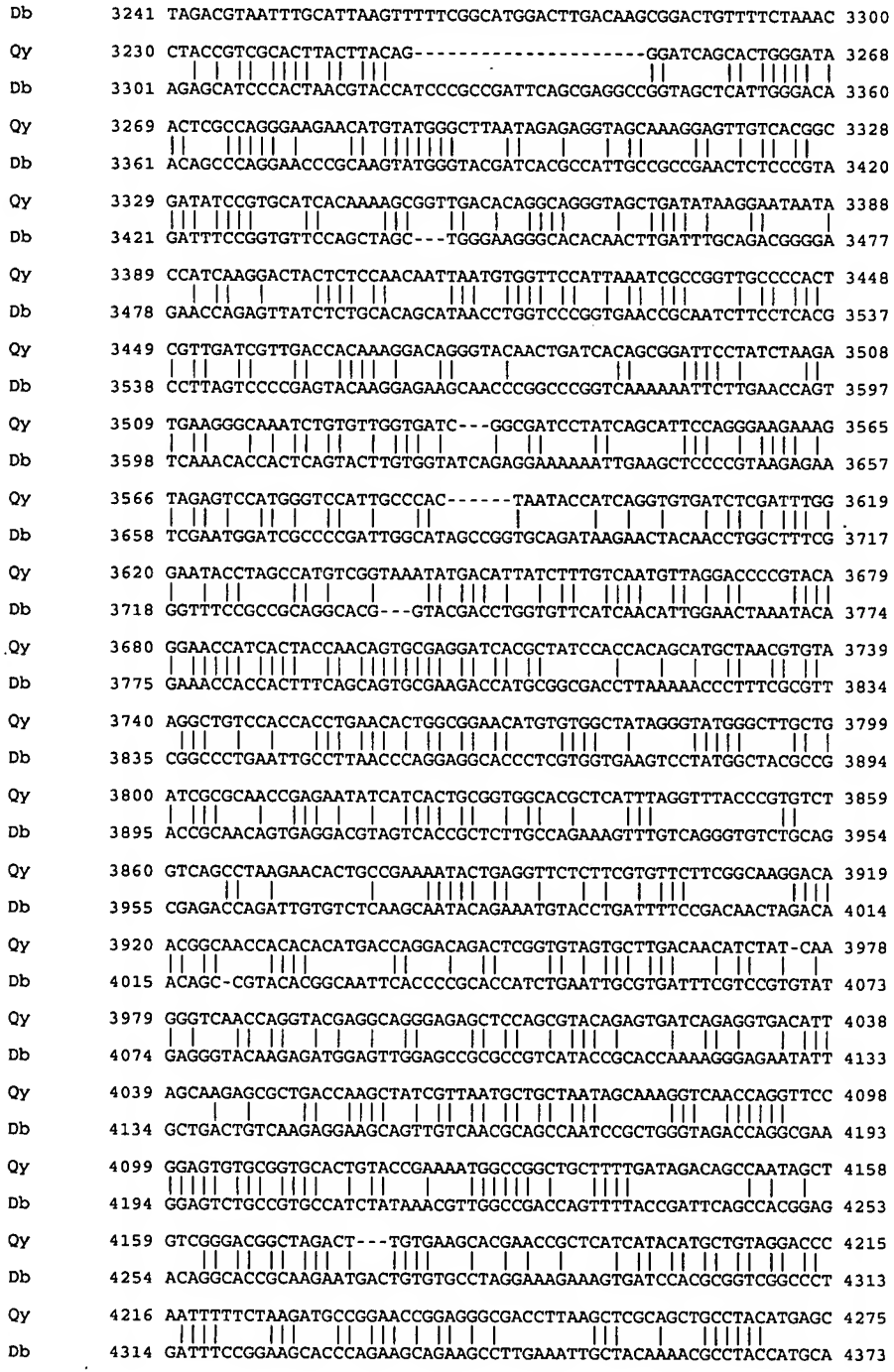
```
Db       4674 AGTACTACAAAAGGAAAATTGTATTCGTACTTCGAAGGCACCAAATTCCATCAAGCAGCA 4733

Qy       4633 AAGGACATTGCCGAAATCCATGCAATGTGGCCCAACAAATCTGAGGCTAATGAGCAGATT 4692

Db       4734 AAAGACATGGCGGAGATAAAGGTCCTGTTCCCTAATGACCAGGAAAGTAATGAACAACTG 4793

Qy       4693 TGCTTGTACATCCTGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGCCCAGTAGAGGAG 4752

Db       4794 TGTGCCTACATATTGGGTGAGACCATGGAAGCAATCCGCGAAAAGTGCCCGGTCGACCAT 4853

Qy       4753 TCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTACGCTATGACGGCT 4812

Db       4854 AACCCGTCGTCTAGCCCGCCCAAAACGTTGCCGTGCCTTTGCATGTATGCCATGACGCCA 4913

Qy       4813 GAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTATGCTCATCATTC 4872

Db       4914 GAAAGGGTCCACAGACTTAGAAGCAATAACGTCAAAGAAGTTACAGTATGCTCCTCCACC 4973

Qy       4873 CTGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGCAAACCAGTCCTG 4932

Db       4974 CCCCTTCCTAAGCACAAAATTAAGAATGTTCAGAAGGTTCAGTGCACGAAAGTAGTCCTG 5033

Qy       4933 TTTTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGCGGAAAT 4982

Db       5034 TTTAATCCGCACACTCCCGCATTCGTTCCCGCCCGTAAGTACATAGAAGT 5083
```

RESULT 11
US-09-507-362-103
; Sequence 103, Application US/09507362
; Publication No. US20030096397A1
; GENERAL INFORMATION:
; APPLICANT: Dubensky Jr., Thomas W.
; Polo, John M.
; Belli, Barbara A.
; Schlesinger, Sondra
; Dryga, Sergey A.
; Frolov, Ilya
; TITLE OF INVENTION: RECOMBINANT ALPHAVIRUS-BASED VECTORS
; WITH REDUCED INHIBITION OF CELLULAR MACRO-MOLECULAR
; SYNTHESIS
; NUMBER OF SEQUENCES: 125
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed Intellectual Property Law Group PLLC
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/507,362
; FILING DATE: 18-Feb-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 930049.457D6 /1196.011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 103:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11740 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 103:
US-09-507-362-103

Query Match 26.0%; Score 1300.4; DB 11; Length 11740;
Best Local Similarity 56.0%; Pred. No. 0;
Matches 2819; Conservative 0; Mismatches 2131; Indels 80; Gaps 16;

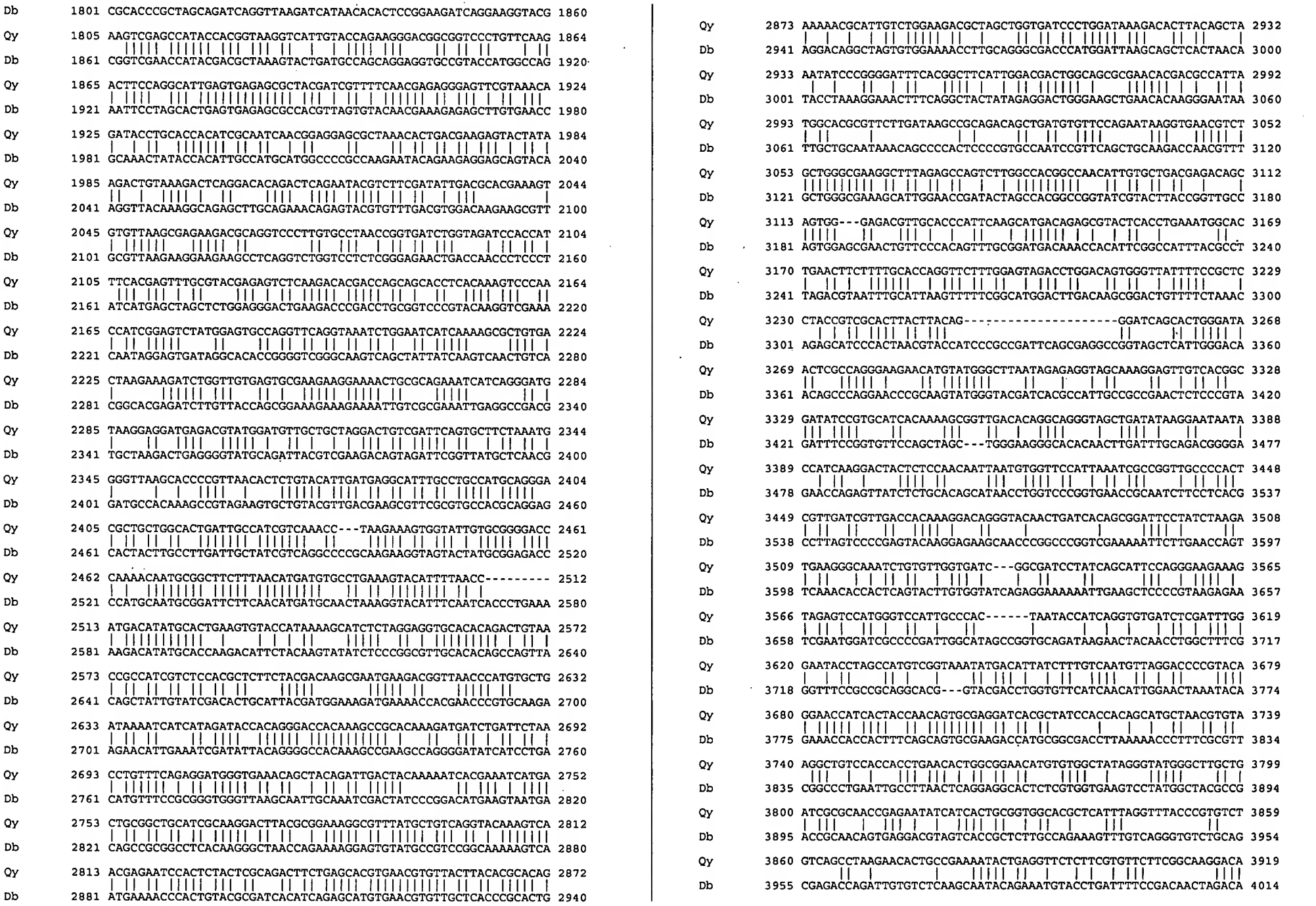
```
Qy         23 ATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGTCGTTACAGC 82

Db         64 AGAAGCCAGTAGTAAACGTAGACGTAGACCCCCAGAGTCCGTTTGTCGTGCAACTGCAAA 123

Qy         83 GGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACCATGCCAATG 142

Db        124 AAAGCTTCCCGCAATTTGAGGTAGTAGCACAGCAGGTCACTCCAAATGACCATGCTAATG 183

Qy        143 CCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACCGGGACCAAG 202

Db        184 CCAGAGCATTTTCGCATCTGGCCAGTAAACTAATCGAGCTGGAGGTTCCTACCACAGCGA 243

Qy        203 TTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACCGCTATCATT 262

Db        244 CGATCTTGGACATAGGCAGCGCACCGGCTCGTAGAATGTTTTCCGAGCACCAGTATCATT 303

Qy        263 GTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGGTATGCAGAAAGAC 322

Db        304 GTGTCTGCCCCATGCGTAGTCCAGAAGACCCGGACCGCATGATGAAATACGCCAGTAAAC 363

Qy        323 TTAAGAAAA------GTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACCTGC 376

Db        364 TGGCGGAAAAAGCGTGCAAGATTACAAACAAGAACTTGCATGAGAAGATTAAGGATCTCC 423

Qy        377 TGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACGCCA 436

Db        424 GGACCGTACTTGATACGCCGGATGCTGAAACACCATCGCTCTGCTTTCACAACGATGTTA 483

Qy        437 CGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCACCGA 496

Db        484 CCTGCAACATGCGTGCCGAATATTCCGTCATGCAGGACGTGTA---TATCAACGCTCCCG 540

Qy        497 CATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTGACA 556

Db        541 GAACTATCTATCATCAGGCTATGAAAGGCGTGCGGACCCTGTACTGGATTGGCTTCGACA 600

Qy        557 CGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACTGGG 616

Db        601 CCACCCAGTTCATGTTCTCGGCTATGGCAGGTTCGTACCCTGCGTACAACACCAACTGGG 660

Qy        617 CTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGGAGA 676

Db        661 CCGACGAGAAAGTCCTTGAAGCGCGTAACATCGGACTTTGCAGCACAAAGCTGAGTGAAG 720

Qy        677 GCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGATCA 736

Db        721 GTAGGACAGGAAAATTGTCGATAATGAGGAAGAAGGAGTTGAAGCCCGGGTCGCGGGTTT 780

Qy        737 TATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGCATC 796

Db        781 ATTTCTCCGTAGGATCGACACTTTATCCAGAACACAGAGCCAGCTTGCAGAGCTGGCATC 840

Qy        797 TTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTG 856

Db        841 TTCCATCGGTGTTCCACTTGAATGGAAAGCAGTCGTACACTTGCCGCTGTGATACAGTGG 900

Qy        857 TCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAG 916

Db        901 TGAGTTGCGAAGGCTACGTAGTGAAGAAAATCACCATCAGTCCCGGGATCACGGGAGAAA 960

Qy        917 TTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATA 976

Db        961 CCGTGGGATACGCGGTTACACACAATAGCGAGGGCTTCTTGCTATGCAAAGTTACTGACA 1020

Qy        977 CGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCG 1036

Db       1021 CAGTAAAAGGAGAACGGGTATCGTTCCCTGTGTGCACGTACATCCCGGCCACCATATGCG 1080
```

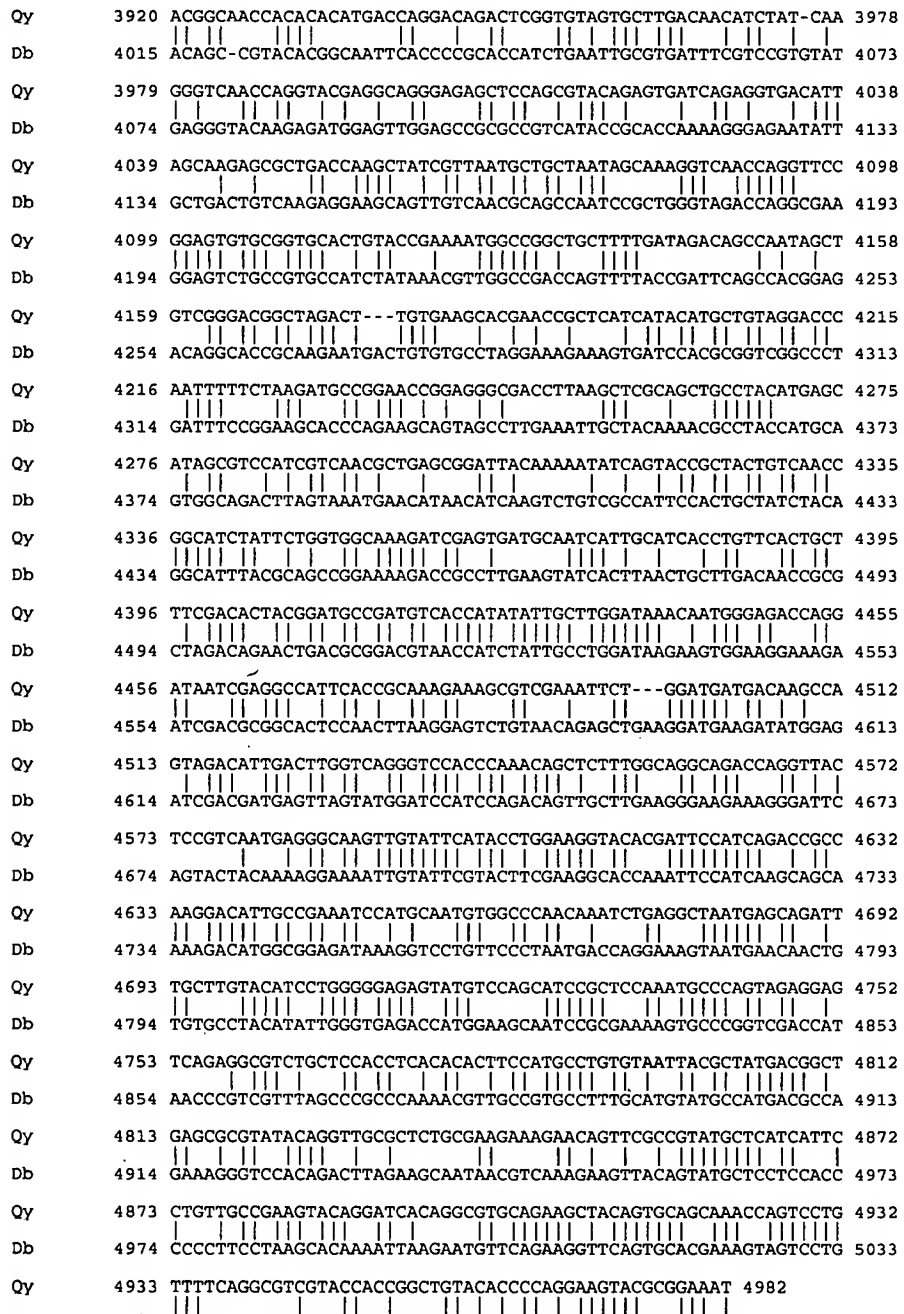




RESULT 12
US-09-507-362-101
; Sequence 101, Application US/09507362
; Publication No. US20030096397A1
; GENERAL INFORMATION:
; APPLICANT: Dubensky Jr., Thomas W.
; Polo, John M.
; Belli, Barbara A.
; Schlesinger, Sondra
; Dryga, Sergey A.
; Frolov, Ilya
; TITLE OF INVENTION: RECOMBINANT ALPHAVIRUS-BASED VECTORS
; WITH REDUCED INHIBITION OF CELLULAR MACRO-MOLECULAR
; SYNTHESIS
; NUMBER OF SEQUENCES: 125
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed Intellectual Property Law Group PLLC
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:

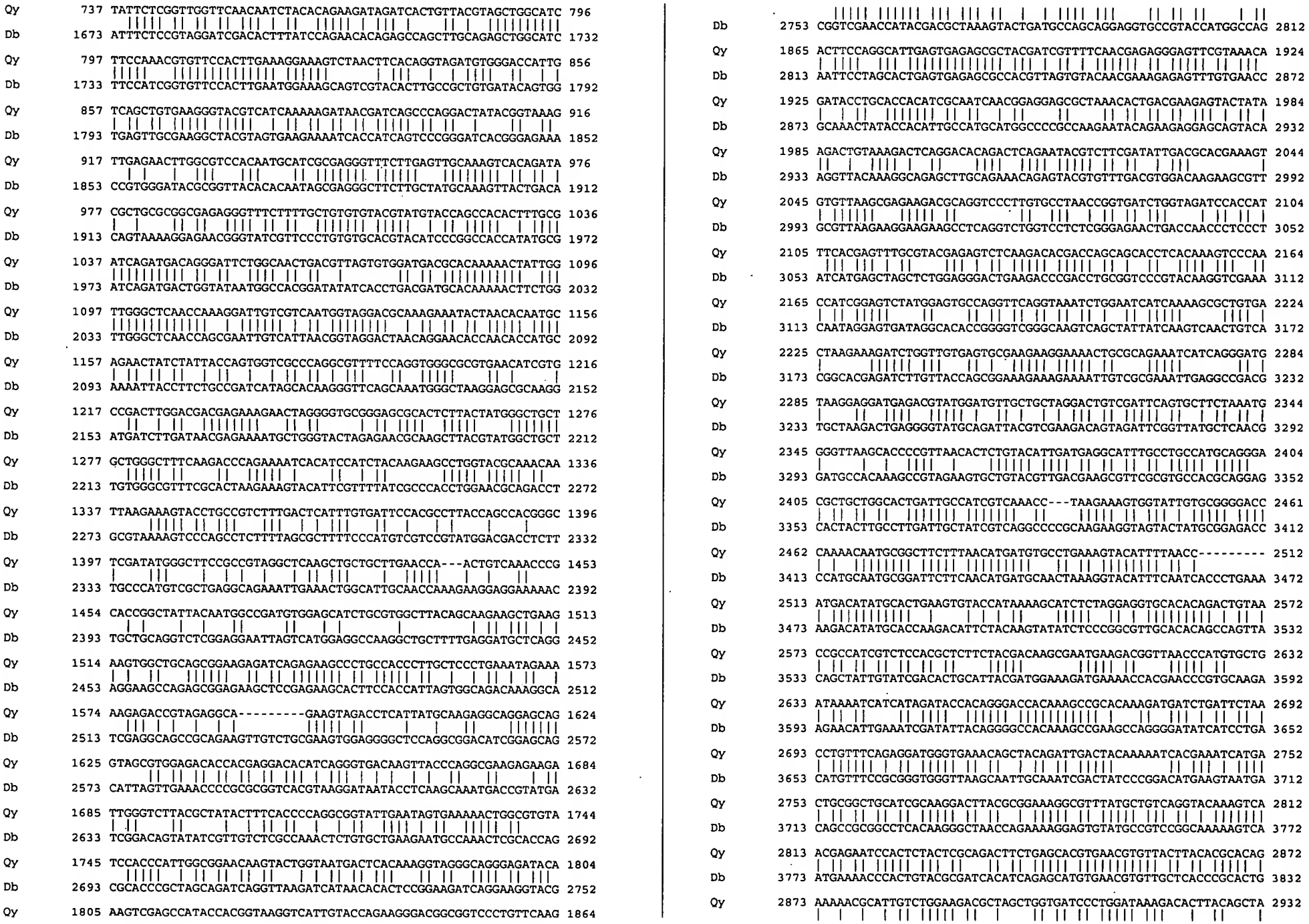
```
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/507,362
;             FILING DATE: 18-Feb-2000
;             CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;             NAME: McMasters, David D.
;             REGISTRATION NUMBER: 33,963
;             REFERENCE/DOCKET NUMBER: 930049.457D6 /1196.011
;       TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (206) 622-4900
;             TELEFAX: (206) 682-6031
;   INFORMATION FOR SEQ ID NO: 101:
;       SEQUENCE CHARACTERISTICS:
;             LENGTH: 8000 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 101:
US-09-507-362-101

  Query Match             25.9%;  Score 1295.6;  DB 11;  Length 8000;
  Best Local Similarity   56.0%;  Pred. No. 0;
  Matches 2816;  Conservative    0;  Mismatches 2134;  Indels   80;  Gaps   16;

Qy         23 ATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGTCGTTACAGC 82
              | | |  |  | |  |||| ||| |||| |  | || |||| |||| |    | |
Db         64 AGAAGCCAGTAGTAAACGTAGACGTAGACCCCCAGAGTCCGTTTGTCGTGCAACTGAAAA 123

Qy         83 GGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACCATGCCAATG 142
                |  || || ||||||||| | | ||||  ||||||||||   ||||||||||| ||||
Db        124 AAAGCTTCCCGCAATTTGAGGTAGTAGCACAGCAGGTCACTCCAAATGACCATGCTAATG 183

Qy        143 CCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACCGGGACCAAG 202
              ||||||| |||||||| |||| |  || || || |||   || ||
Db        184 CCAGAGCATTTTCGCATCTGGCCAGTAAACTAATCGAGCTGGAGGTTCCTACCACAGCGA 243

Qy        203 TTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACCGCTATCATT 262
              |||||||||||  || || || || |   |       |||| | ||||  |||||||
Db        244 CGATCTTGGACATAGGCAGCGCACCGGCTCGTAGAATGTTTTCCGAGCACCAGTATCATT 303

Qy        263 GTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGGTATGCAGAAAGAC 322
              || ||||||| |||   ||  | ||||||||||||| |  |       || ||  . | ||
Db        304 GTGTCTGCCCCATGCGTAGTCCAGAAGACCCGGACCGCATGATGAAATACGCCAGTAAAC 363

Qy        323 TTAAGAAAA------GTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACCTGC 376
              |    | |||      |  | |||||  ||||||| |       |||     || || |
Db        364 TGGCGGAAAAAGCGTGCAAGATTACAAACAAGAACTTGCATGAGAAGATTAAGGATCTCC 423

Qy        377 TGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACGCCA 436
              |   ||  |  || || || || || || ||||| || ||  | ||||  || | |
Db        424 GGACCGTACTTGATACGCCGGATGCTGAAACACCATCGCTCTGCTTTCACAACGATGTTA 483

Qy        437 CGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCACCGA 496
              | || |       ||         | ||    || || |||||     || | || ||
Db        484 CCTGCAACATGCGTGCCGAATATTCCGTCATGCAGGACGTGTA---TATCAACGCTCCCG 540

Qy        497 CATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTGACA 556
              | | ||||| || |||||  | ||||| ||  ||||  | |||||||| ||||| ||||
Db        541 GAACTATCTATCATCAGGCTATGAAAGGCGTGCGGACCCTGTACTGGATTGGCTTCGACA 600

Qy        557 CGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACTGGG 616
              | ||||  || |||| |      ||||||||||| |||||| | |||||||| |||||||
Db        601 CCACCCAGTTCATGTTCTCGGCTATGGCAGGTTCGTACCCTGCGTACAACACCAACTGGG 660

Qy        617 CTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGGAGA 676
              | ||||||| |||  | ||||| |||||||| || ||  | | || | ||    ||
Db        661 CCGACGAGAAAGTCCTTGAAGCGCGTAACATCGGACTTTGCAGCACAAAGCTGAGTGAAG 720

Qy        677 GCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGATCA 736

                | |||   |||||| | || ||  | |||||||||  | |  | ||          | |
Db        721 GTAGGACAGGAAAATTGTCGATAATGAGGAAGAAGGAGTTGAAGCCCGGGTCGCGGGTTT 780

Qy        737 TATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGCATC 796
              ||||| || || || ||| | ||  ||||| | ||| |     || |  ||||||||||
Db        781 ATTTCTCCGTAGGATCGACACTTTATCCAGAACACAGAGCCAGCTTGCAGAGCTGGCATC 840

Qy        797 TTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTG 856
              |||||    |||||||||||||| ||||||       | |||  |  | ||||  ||  | |
Db        841 TTCCATCGGTGTTCCACTTGAATGGAAAGCAGTCGTACACTTGCCGCTGTGATACAGTGG 900

Qy        857 TCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAG 916
              | || || ||||| |||||  | || || || || ||||| || ||  |     ||  ||
Db        901 TGAGTTGCGAAGGCTACGTAGTGAAGAAAATCACCATCAGTCCCGGGATCACGGGAGAAA 960

Qy        917 TTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATA 976
              | |   | |||   |||     || ||||||| ||||||   |||||||| || || |
Db        961 CCGTGGGATACGCGGTTACACACAATAGCGAGGGCTTCTTGCTATGCAAAGTTACTGACA 1020

Qy        977 CGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCG 1036
              | |    || ||  |||| || ||  |||||| |||||  | || |||||  | ||||
Db       1021 CAGTAAAAGGAGAACGGGTATCGTTCCCTGTGTGCACGTACATCCCGGCCACCATATGCG 1080

Qy       1037 ATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTATTGG 1096
              |||||||||| || ||  |||| || ||  |        || || ||||||||||  |||
Db       1081 ATCAGATGACTGGTATAATGGCCACGGATATATCACCTGACGATGCACAAAAACTTCTGG 1140

Qy       1097 TTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAATGC 1156
              |||||||||||  | |||||| | || |||||||||  | || || || ||||| ||||
Db       1141 TTGGGCTCAACCAGCGAATTGTCATTAACGGTAGGACTAACAGGAACACCAACACCATGC 1200

Qy       1157 AGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATCGTG 1216
              | || || ||  | ||  |  | || || | |||   ||  |||||     || |     |
Db       1201 AAAATTACCTTCTGCCGATCATAGCACAAGGGTTCAGCAAATGGGCTAAGGAGCGCAAGG 1260

Qy       1217 CCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCTGCT 1276
                || | ||  ||||||||   || ||     | || ||||  |||||    |||||||
Db       1261 ATGATCTTGATAACGAGAAAATGCTGGGTACTAGAGAACGCAAGCTTACGTATGGCTGCT 1320

Qy       1277 GCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCTGGTACGCAAACAA 1336
              |||||  ||    ||  |||||  |    || | ||        |||||  |||||  ||
Db       1321 TGTGGGCGTTTCGCACTAAGAAAGTACATTCGTTTTATCGCCCACCTGGAACGCAGACCT 1380

Qy       1337 TTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCACGGGC 1396
              |||||  || |||   |||  | | |||   ||  |   | |   |   |
Db       1381 GCGTAAAAGTCCCAGCCTCTTTTAGCGCTTTTCCCATGTCGTCCGTATGGACGACCTCTT 1440

Qy       1397 TCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCA---ACTGTCAAACCCG 1453
              |    |||    | | |  |  | || |||    |  |||||   |  |   ||
Db       1441 TGCCCATGTCGCTGAGGCAGAAATTGAAACTGGCATTGCAACCAAAGAAGGAGGAAAAAC 1500

Qy       1454 CACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAG 1513
              | |    |  |   ||   ||  | |                  | ||| ||| | |
Db       1501 TGCTGCAGGTCTCGGAGGAATTAGTCATGGAGGCCAAGGCTGCTTTTGAGGATGCTCAGG 1560

Qy       1514 AAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAA 1573
              | |  ||   |||||  || || ||||||| || ||||| ||  |  | || | ||  |
Db       1561 AGGAAGCCAGAGCGGAGAAGCTCCGAGAAGCACTTCCACCATTAGTGGCAGACAAAGGCA 1620

Qy       1574 AAGAGACCGTAGAGGCA---------GAAGTAGACCTCATTATGCAAGAGGCAGGAGCAG 1624
              |||  |  |  | |          ||||| ||     |   ||    |||||||
Db       1621 TCGAGGCAGCCGCAGAAGTTGTCTGCGAAGTGGAGGGGCTCCAGGCGGACATCGGAGCAG 1680

Qy       1625 GTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAGA 1684
              || || || || || || ||| | ||| | | | | ||    ||      | ||
Db       1681 CATTAGTTGAAACCCCGCGCGGTCACGTAAGGATAATACCTCAAGCAAATGACCGTATGA 1740

Qy       1685 TTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGTA 1744
              | ||   ||    |  | || ||  |  | ||| |||| | ||  ||||| ||
Db       1741 TCGGACAGTATATCGTTGTCTCGCCAAACTCTGTACTGAAGAATGCCAAACTCGCACCAG 1800

Qy       1745 TCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATACA 1804
              |||||  | || || || ||   | | || || |||   || ||  |||| || |||
```

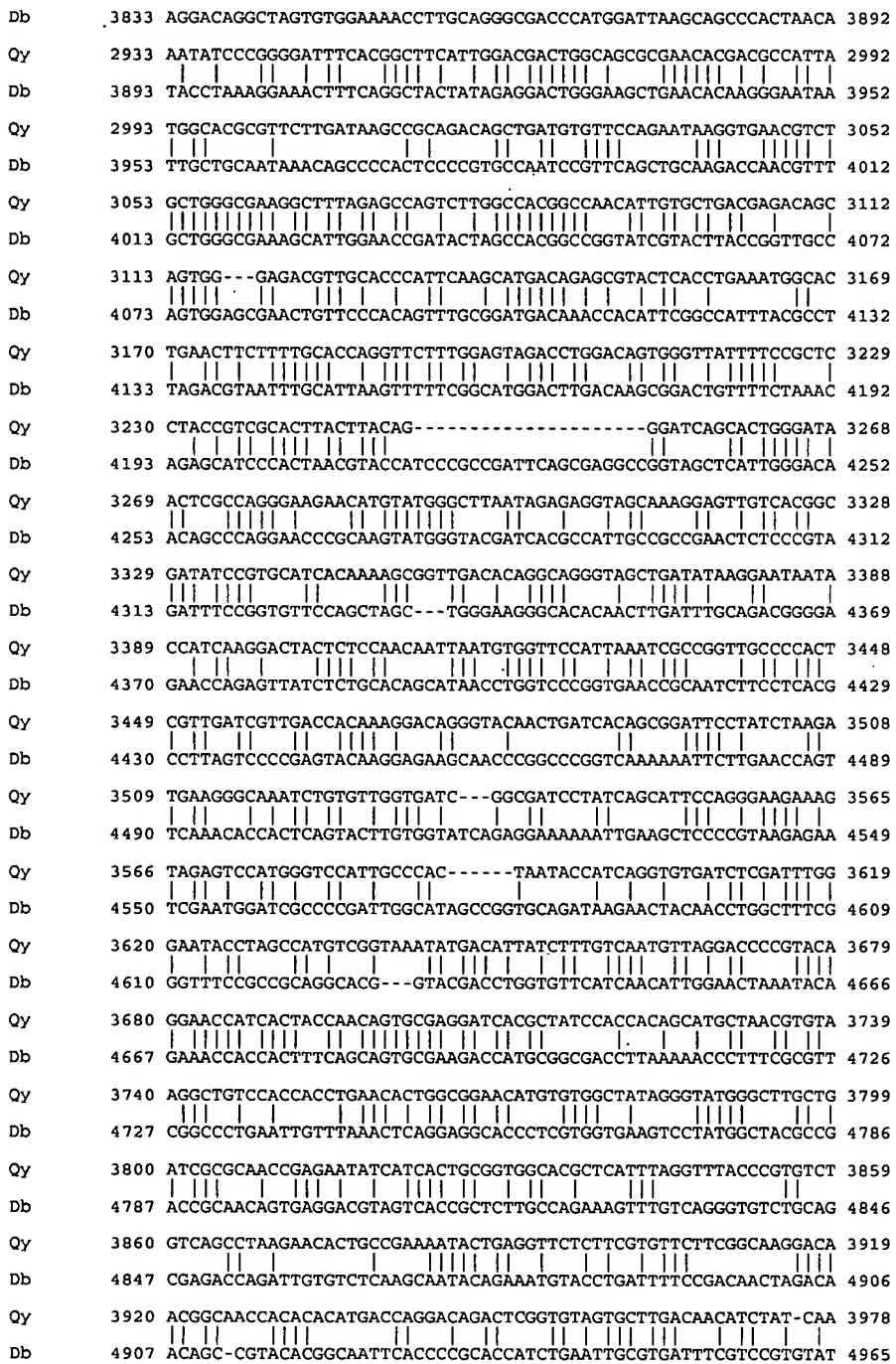


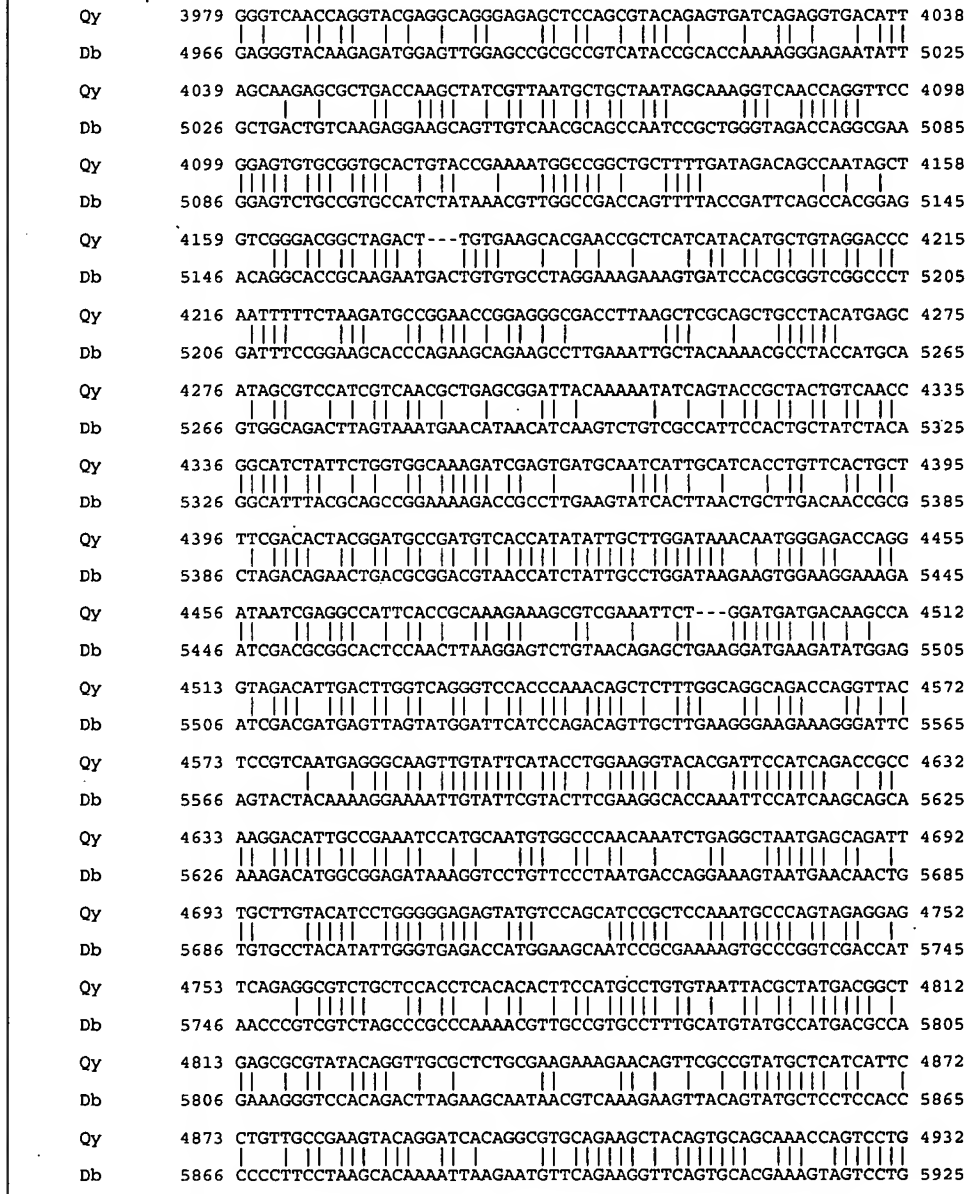


```
Qy       3920 ACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTGCTTGACAACATCTAT-CAA 3978
Db       4015 ACAGC-CGTACACGGCAATTCACCCCGCACCATCTGAATTGCGTGATTTCGTCCGTGTAT 4073

Qy       3979 GGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGTACAGAGTGATCAGAGGTGACATT 4038
Db       4074 GAGGGTACAAGAGATGGAGTTGGAGCCGCGCCGTCATACCGCACCAAAAGGGAGAATATT 4133

Qy       4039 AGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGCAAAGGTCAACCAGGTTCC 4098
Db       4134 GCTGACTGTCAAGAGGAAGCAGTTGTCAACGCAGCCAATCCGCTGGGTAGACCAGGCGAA 4193

Qy       4099 GGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTTGATAGACAGCCAATAGCT 4158
Db       4194 GGAGTCTGCCGTGCCATCTATAAACGTTGGCCGACCAGTTTTACCGATTCAGCCACGGAG 4253

Qy       4159 GTCGGGACGGCTAGACT---TGTGAAGCACGAACCGCTCATCATACATGCTGTAGGACCC 4215
Db       4254 ACAGGCACCGCAAGAATGACTGTGTGCCTAGGAAAGAAAGTGATCCACGCGGTCGGCCCT 4313

Qy       4216 AATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTAAGCTCGCAGCTGCCTACATGAGC 4275
Db       4314 GATTTCCGGAAGCACCCAGAAGCAGTAGCCTTGAAATTGCTACAAAACGCCTACCATGCA 4373

Qy       4276 ATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTACCGCTACTGTCAACC 4335
Db       4374 GTGGCAGACTTAGTAAATGAACATAACATCAAGTCTGTCGCCATTCCACTGCTATCTACA 4433

Qy       4336 GGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCATCACCTGTTCACTGCT 4395
Db       4434 GGCATTTACGCAGCCGGAAAAGACCGCCTTGAAGTATCACTTAACTGCTTGACAACCGCG 4493

Qy       4396 TTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAACAATGGGAGACCAGG 4455
Db       4494 CTAGACAGAACTGACGCGGACGTAACCATCTATTGCCTGGATAAGAAGTGGAAGGAAAGA 4553

Qy       4456 ATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATTCT---GGATGATGACAAGCCA 4512
Db       4554 ATCGACGCGGCACTCCAACTTAAGGAGTCTGTAACAGAGCTGAAGGATGAAGATATGGAG 4613

Qy       4513 GTAGACATTGACTTGGTCAGGGTCCACCCAAACAGCTCTTTGGCAGGCAGACCAGGTTAC 4572
Db       4614 ATCGACGATGAGTTAGTATGGATCCATCCAGACAGTTGCTTGAAGGGAAGAAAGGGATTC 4673

Qy       4573 TCCGTCAATGAGGGCAAGTTGTATTCATACCTGGAAGGTACACGATTCCATCAGACCGCC 4632
Db       4674 AGTACTACAAAAGGAAAATTGTATTCGTACTTCGAAGGCACCAAATTCCATCAAGCAGCA 4733

Qy       4633 AAGGACATTGCCGAAATCCATGCAATGTGGCCCAACAAATCTGAGGCTAATGAGCAGATT 4692
Db       4734 AAAGACATGGCGGAGATAAAGGTCCTGTTCCCTAATGACCAGGAAAGTAATGAACAACTG 4793

Qy       4693 TGCTTGTACATCCTGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGCCCAGTAGAGGAG 4752
Db       4794 TGTGCCTACATATTGGGTGAGACCATGGAAGCAATCCGCGAAAAGTGCCCGGTCGACCAT 4853

Qy       4753 TCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTACGCTATGACGGCT 4812
Db       4854 AACCCGTCGTTTAGCCCGCCCAAAACGTTGCCGTGCCTTTGCATGTATGCCATGACGCCA 4913

Qy       4813 GAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTATGCTCATCATTC 4872
Db       4914 GAAAGGGTCCACAGACTTAGAAGCAATAACGTCAAAGAAGTTACAGTATGCTCCTCCACC 4973

Qy       4873 CTGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGCAAACCAGTCCTG 4932
Db       4974 CCCCTTCCTAAGCACAAAATTAAGAATGTTCAGAAGGTTCAGTGCACGAAAGTAGTCCTG 5033

Qy       4933 TTTTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGCGGAAAT 4982
Db       5034 TTTAATCCGCACACTCCCGCATTCGTTCCCGCCCGTAAGTACATAGAAGT 5083
```

```
RESULT 13
US-09-733-042-1
; Sequence 1, Application US/09733042
; Patent No. US20020168709A1
; GENERAL INFORMATION:
;  APPLICANT:   Hennecke, Frank
;  APPLICANT:  Renner, Wolfgang A.
;  TITLE OF INVENTION:   Replicon Based Activation of Endogenous Genes
;  FILE REFERENCE:   1700.0100001
;  CURRENT APPLICATION NUMBER: US/09/733,042
;  CURRENT FILING DATE:  2000-12-11
;  PRIOR APPLICATION NUMBER:   US 60/169,988
;  PRIOR FILING DATE:   1999-12-10
;  NUMBER OF SEQ ID NOS:   49
;  SOFTWARE:   PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 11282
;   TYPE: DNA
;   ORGANISM: pCYTts
US-09-733-042-1

  Query Match           25.9%;  Score 1295.6;  DB 10;  Length 11282;
  Best Local Similarity   56.0%;  Pred. No. 0;
  Matches 2816;  Conservative    0;  Mismatches 2134;  Indels   80;  Gaps   16;

Qy         23 ATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGTCGTTACAGC 82
Db        956 AGAAGCCAGTAGTAAACGTAGACGTAGACCCCCAGAGTCCGTTTGTCGTGCAACTGCAAA 1015

Qy         83 GGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACCATGCCAATG 142
Db       1016 AAAGCTTCCCGCAATTTGAGGTAGTAGCACAGCAGGTCACTCCAAATGACCATGCTAATG 1075

Qy        143 CCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACCGGGACCAAG 202
Db       1076 CCAGAGCATTTTCGCATCTGGCCAGTAAACTAATCGAGCTGGAGGTTCCTACCACAGCGA 1135

Qy        203 TTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACCGCTATCATT 262
Db       1136 CGATCTTGGACATAGGCAGCGCACCGGCTCGTAGAATGTTTTCCGAGCACCAGTATCATT 1195

Qy        263 GTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGGTATGCAGAAAGAC 322
Db       1196 GTGTCTGCCCCATGCGTAGTCCAGAAGACCCGGACCGCATGATGAAATACGCCAGTAAAC 1255

Qy        323 TTAAGAAAA------GTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACCTGC 376
Db       1256 TGGCGGAAAAAGCGTGCAAGATTACAAACAAGAACTTGCATGAGAAGATTAAGGATCTCC 1315

Qy        377 TGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACGCCA 436
Db       1316 GGACCGTACTTGATACGCCGGATGCTGAAACACCATCGCTCTGCTTTCACAACGATGTTA 1375

Qy        437 CGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCACCGA 496
Db       1376 CCTGCAACATGCGTGCCGAATATTCCGTCATGCAGGACGTGTA---TATCAACGCTCCCG 1432

Qy        497 CATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTGACA 556
Db       1433 GAACTATCTATCATCAGGCTATGAAAGGCGTGCGGACCCTGTACTGGATTGGCTTCGACA 1492

Qy        557 CGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACTGGG 616
Db       1493 CCACCCAGTTCATGTTCTCGGCTATGGCAGGTTCGTACCCTGCGTACAACACCAACTGGG 1552

Qy        617 CTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGGAGA 676
Db       1553 CCGACGAGAAAGTCCTTGAAGCGCGTAACATCGGACTTTGCAGCACAAAGCTGAGTGAAG 1612

Qy        677 GCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGATCA 736
Db       1613 GTAGGACAGGAAAATTGTCGATAATGAGGAAGAAGGAGTTGAAGCCCGGGTCGCGGGTTT 1672
```



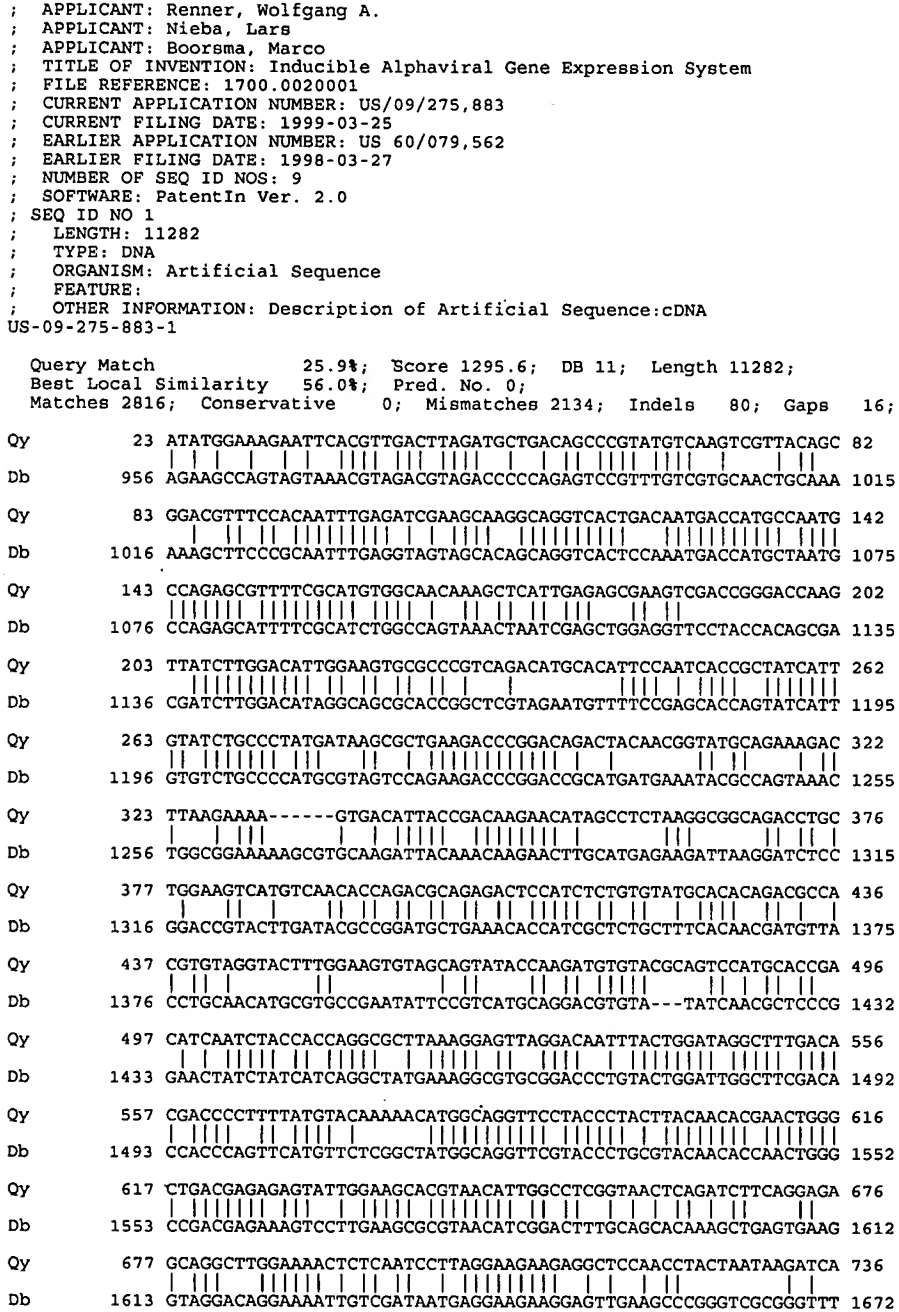


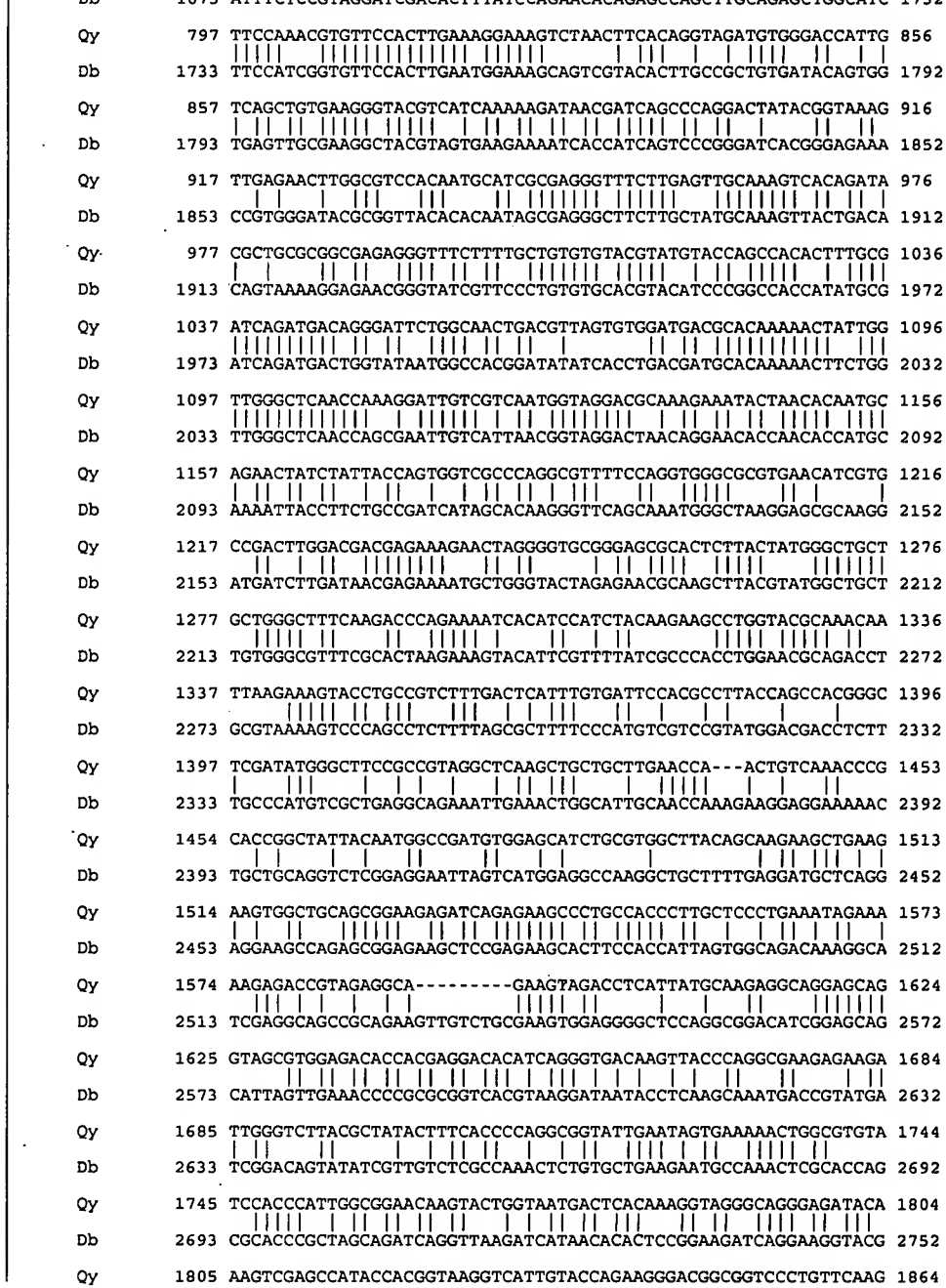


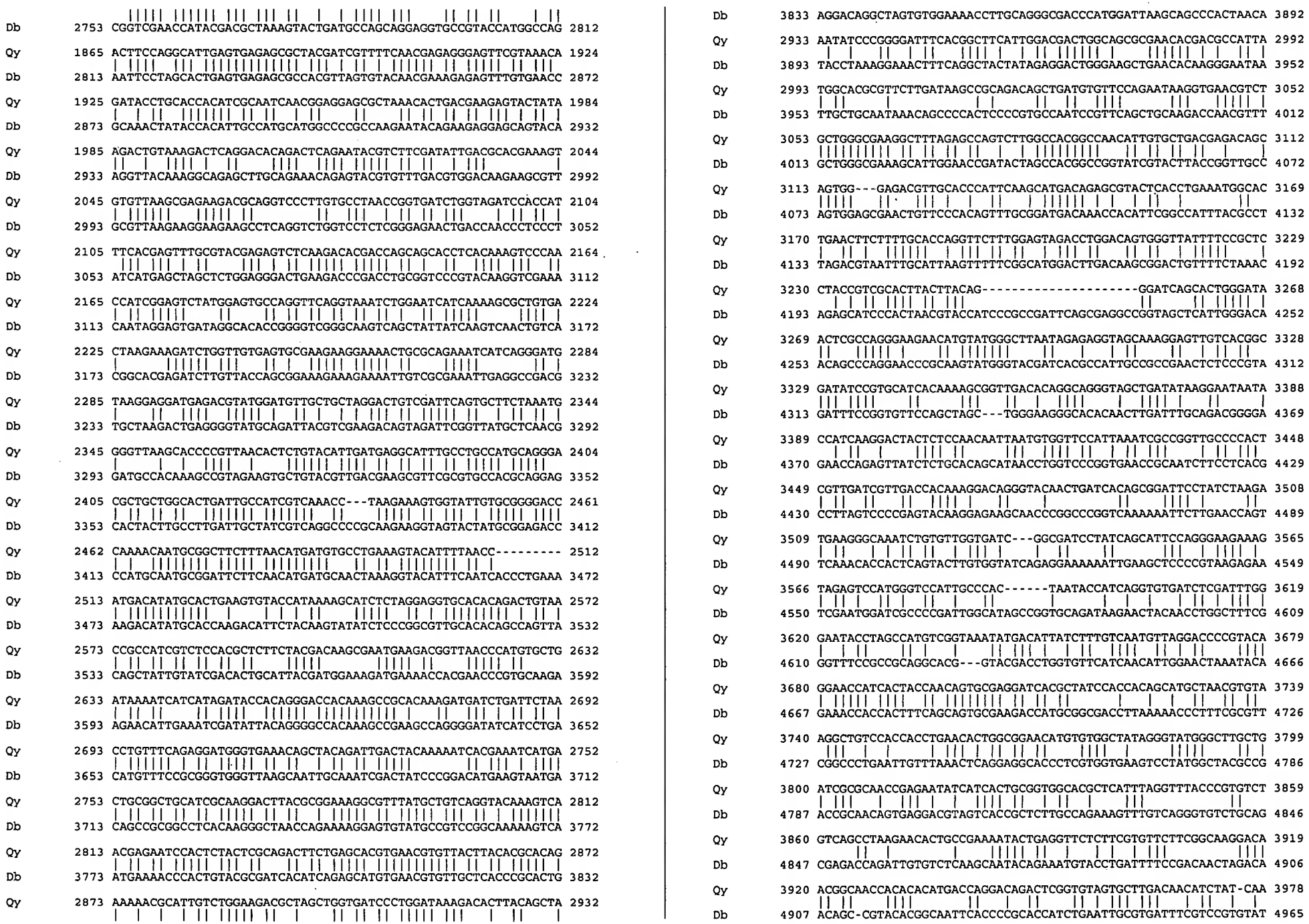
RESULT 14
US-09-275-883-1
; Sequence 1, Application US/09275883

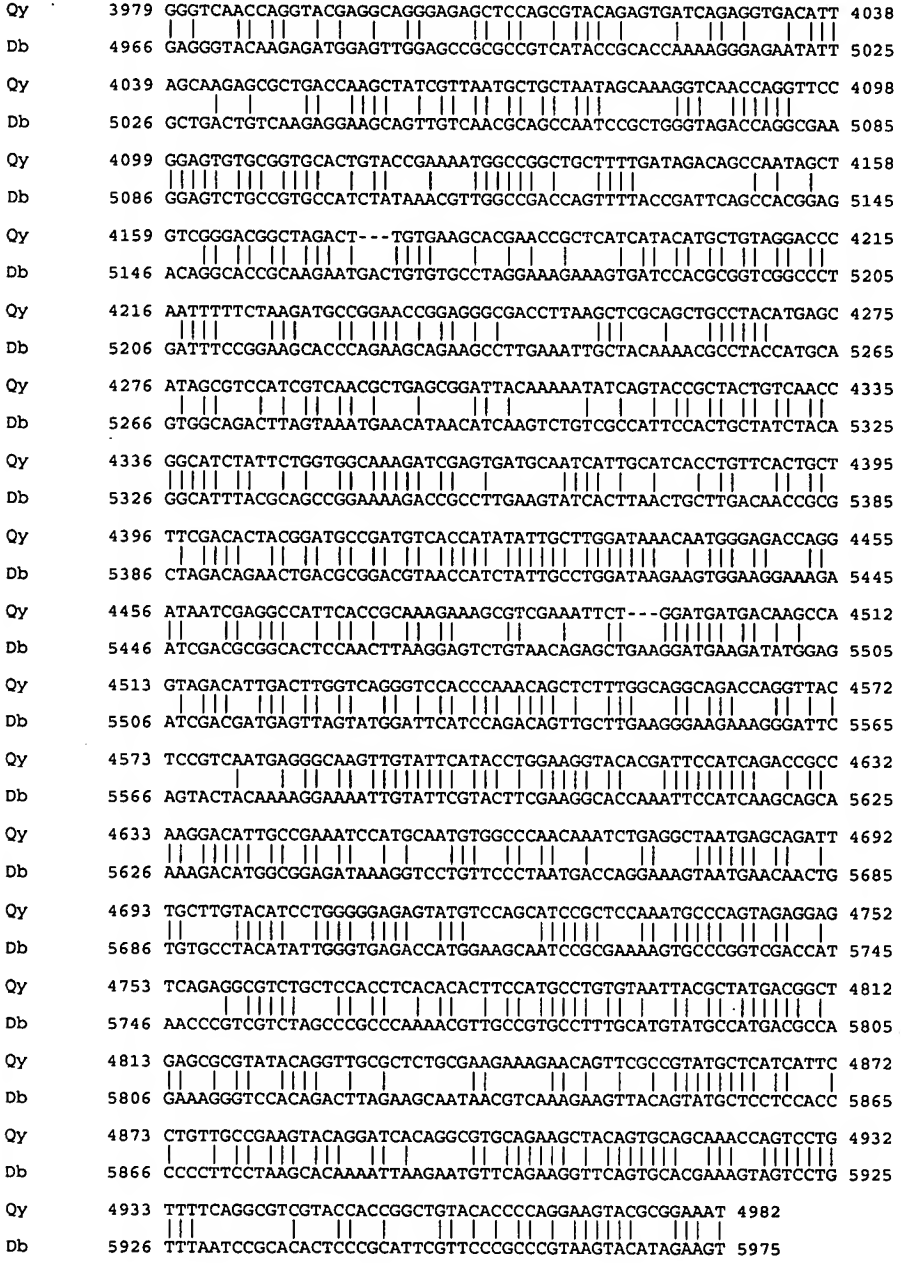
; Publication No. US20030053988A1
; GENERAL INFORMATION:
; APPLICANT: Renner, Wolfgang A.
; APPLICANT: Nieba, Lars
; APPLICANT: Boorsma, Marco
; TITLE OF INVENTION: Inducible Alphaviral Gene Expression System
; FILE REFERENCE: 1700.0020001
; CURRENT APPLICATION NUMBER: US/09/275,883
; CURRENT FILING DATE: 1999-03-25
; EARLIER APPLICATION NUMBER: US 60/079,562
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 11282
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:cDNA
US-09-275-883-1

Query Match 25.9%; Score 1295.6; DB 11; Length 11282;
Best Local Similarity 56.0%; Pred. No. 0;
Matches 2816; Conservative 0; Mismatches 2134; Indels 80; Gaps 16;





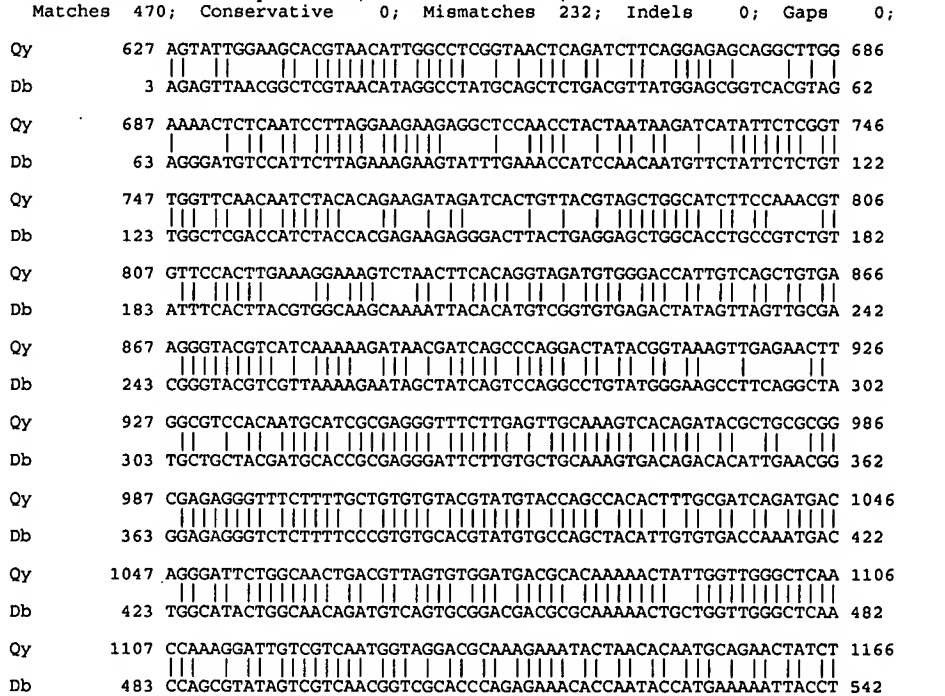




RESULT 15
US-10-123-101-51
; Sequence 51, Application US/10123101

; Publication No. US20030148262A1
; GENERAL INFORMATION:
; APPLICANT: POLO, John
; APPLICANT: PERRI, Silvia
; APPLICANT: THUDIUM, Kent
; TITLE OF INVENTION: CHIMERIC ALPHAVIRUS REPLICON PARTICLES
; FILE REFERENCE: 2300-17924 / PP17924.002
; CURRENT APPLICATION NUMBER: US/10/123,101
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: 60/295,451
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
; LENGTH: 705
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: VEE nsP fragment oligo 1
US-10-123-101-51

Query Match 6.6%; Score 330.8; DB 12; Length 705;
Best Local Similarity 67.0%; Pred. No. 1e-97;
Matches 470; Conservative 0; Mismatches 232; Indels 0; Gaps 0;



Search completed: November 16, 2003, 00:28:09
Job time : 1549 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 12:07:42 ; Search time 308 Seconds
(without alignments)
7165.311 Million cell updates/sec

Title: US-10-023-649-1_COPY_1_5000
Perfect score: 5000
Sequence: 1 accctacaaactaatcgatc.....ataattctagaaacgccacc 5000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

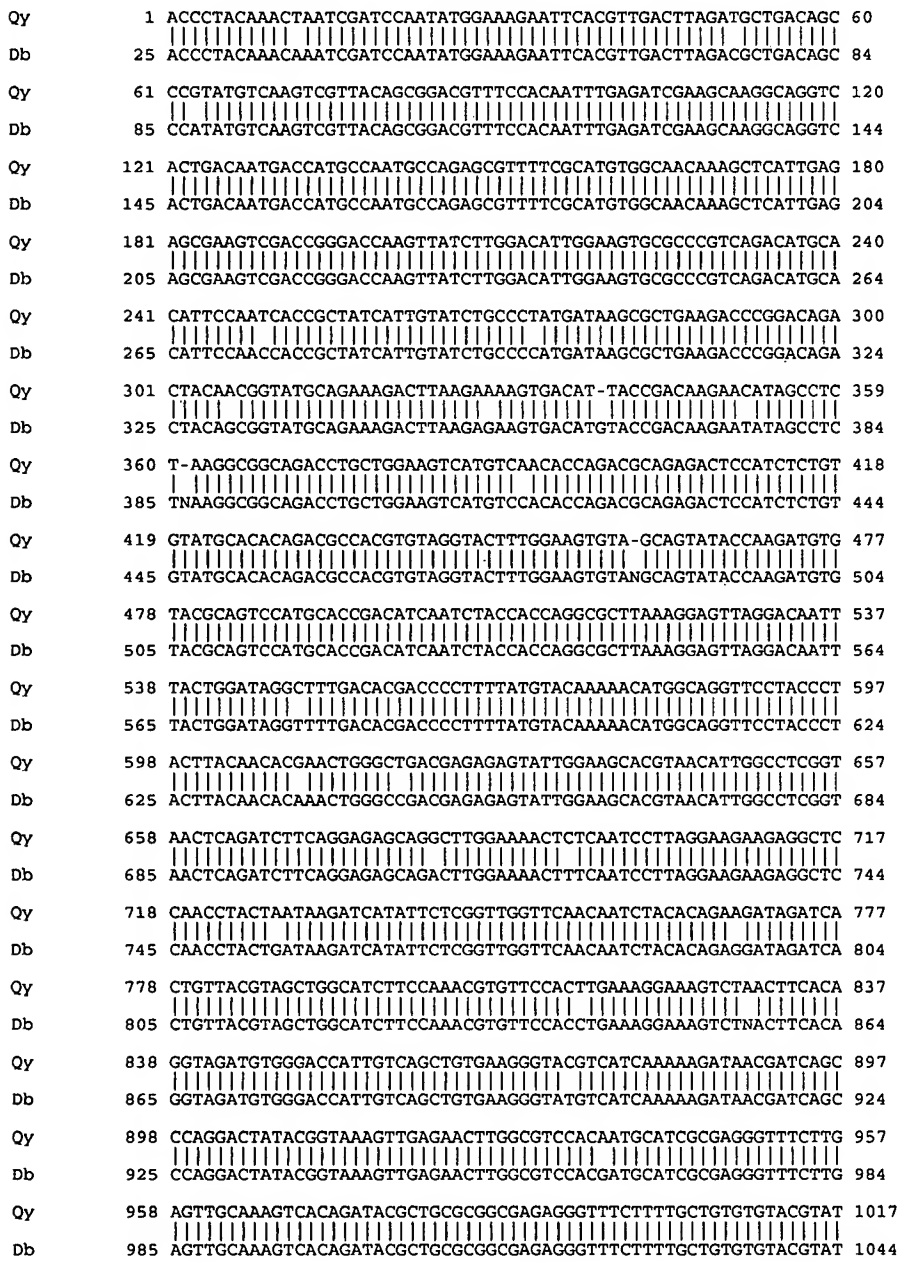
SUMMARIES

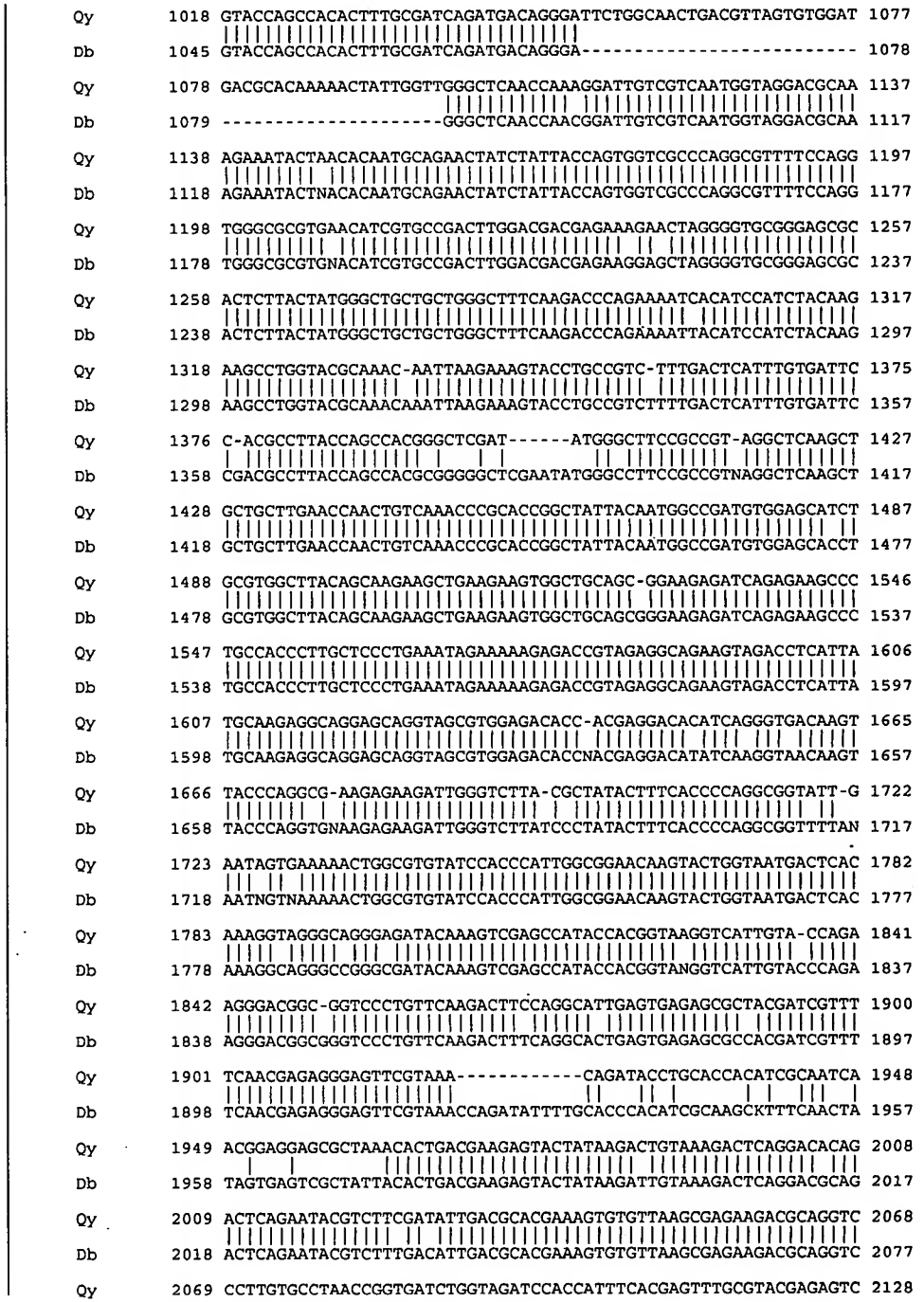
Result No.	Score	% Query Match	Length	DB	ID	Description
1	4492	89.8	11492	3	US-08-991-840A-1	Sequence 1, Appli
2	2118	42.4	11459	3	US-09-454-721A-3	Sequence 3, Appli
3	2074.8	41.5	11464	3	US-08-991-840A-2	Sequence 2, Appli
4	1564	31.3	11517	1	US-07-920-281C-1	Sequence 1, Appli
5	1564	31.3	11517	3	US-08-466-277-1	Sequence 1, Appli
6	1564	31.3	15538	4	US-09-554-337-1	Sequence 1, Appli
7	1473.8	29.5	8100	4	US-09-554-337-4	Sequence 4, Appli
8	1333.2	26.7	11663	1	US-08-446-932-1	Sequence 1, Appli
9	1333.2	26.7	11663	1	US-08-801-263A-1	Sequence 1, Appli
10	1333.2	26.7	11663	1	US-08-801-263A-7	Sequence 7, Appli
11	1333.2	26.7	11663	3	US-09-102-248-1	Sequence 1, Appli
12	1333.2	26.7	11663	3	US-09-102-248-7	Sequence 7, Appli
13	1333.2	26.7	11663	4	US-09-367-764-1	Sequence 1, Appli
14	1333.2	26.7	11663	4	US-09-367-764-7	Sequence 7, Appli
15	1303.6	26.1	11703	1	US-08-801-263A-8	Sequence 8, Appli
16	1303.6	26.1	11703	3	US-09-102-248-8	Sequence 8, Appli
17	1303.6	26.1	11703	4	US-09-367-764-8	Sequence 8, Appli
18	1302.2	26.0	11717	1	US-08-801-263A-4	Sequence 4, Appli
19	1302.2	26.0	11717	3	US-09-102-248-4	Sequence 4, Appli
20	1302.2	26.0	11717	4	US-09-367-764-4	Sequence 4, Appli
21	1300.4	26.0	8000	4	US-09-415-784-102	Sequence 102, App
22	1300.4	26.0	8000	4	US-09-415-785A-102	Sequence 102, App
23	1300.4	26.0	8000	4	US-08-944-465-102	Sequence 102, App
24	1300.4	26.0	8000	4	US-09-415-868-102	Sequence 102, App
25	1300.4	26.0	8000	4	US-09-415-900-102	Sequence 102, App
26	1300.4	26.0	11740	4	US-09-415-784-103	Sequence 103, App
27	1300.4	26.0	11740	4	US-09-415-785A-103	Sequence 103, App
28	1300.4	26.0	11740	4	US-08-944-465-103	Sequence 103, App
29	1300.4	26.0	11740	4	US-09-415-868-103	Sequence 103, App
30	1300.4	26.0	11740	4	US-09-415-900-103	Sequence 103, App
31	1298.8	26.0	9951	4	US-09-193-707-3	Sequence 3, Appli
32	1298.8	26.0	10524	4	US-09-193-707-4	Sequence 4, Appli
33	1298.8	26.0	11927	4	US-09-193-707-5	Sequence 5, Appli
34	1298.8	26.0	13905	3	US-08-972-218-1	Sequence 1, Appli
35	1298.8	26.0	13905	4	US-09-193-707-1	Sequence 1, Appli
36	1295.6	25.9	8000	4	US-09-415-784-101	Sequence 101, App
37	1295.6	25.9	8000	4	US-09-415-785A-101	Sequence 101, App
38	1295.6	25.9	8000	4	US-08-944-465-101	Sequence 101, App
39	1295.6	25.9	8000	4	US-09-415-868-101	Sequence 101, App
40	1295.6	25.9	8000	4	US-09-415-900-101	Sequence 101, App
41	1295.6	25.9	11282	4	US-09-733-042-1	Sequence 1, Appli
42	1294	25.9	16656	1	US-08-741-881-1	Sequence 1, Appli
43	1294	25.9	16656	1	US-08-739-158-1	Sequence 1, Appli
44	1294	25.9	16656	2	US-08-739-167-1	Sequence 1, Appli
45	1294	25.9	16656	3	US-08-404-796-1	Sequence 1, Appli

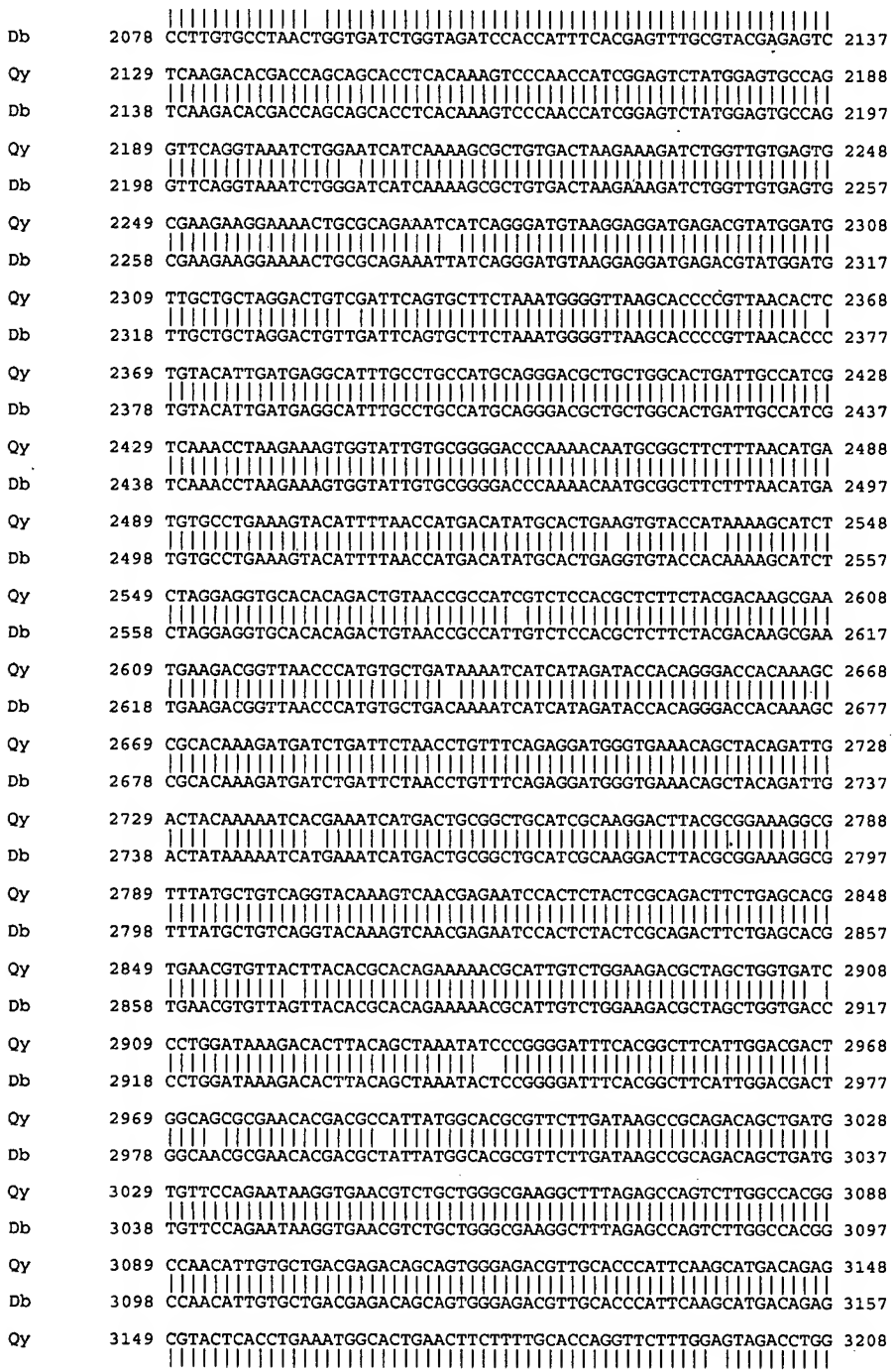
ALIGNMENTS

```
RESULT 1
US-08-991-840A-1
; Sequence 1, Application US/08991840A
; Patent No. 6261570
;  GENERAL INFORMATION:
;    APPLICANT:  Michael D. Parker
;    APPLICANT:  Jonathan F. Smith
;    APPLICANT:  Bruce Crise
;    APPLICANT:  Mark Steve Oberste
;    APPLICANT:  Shannon Schmura
;    TITLE OF INVENTION:  Live Attenuated Virus Vaccines for Eastern Equine Encephalitis
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Charles H. Harris
;      STREET:  USA MRMC -  MRMC-JA
;      CITY:  FORT DETRICK, FREDERICK
;      STATE:  MARYLAND
;      COUNTRY:  USA
;      ZIP:  21702-5012
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh 7.5
;      SOFTWARE:  Microsoft Word 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/991,840A
;      FILING DATE:  December 16, 1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  Provisional Application 60/047162,
;      FILING DATE:  May 20, 1997
;      APPLICATION NUMBER:  Provisional Application 60/053,652
;      FILING DATE:  July 24, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles H. Harris
;      REGISTRATION NUMBER:  34,616
;      REFERENCE/DOCKET NUMBER:  003/058/SAP RIID 96-01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 619-2065
;      TELEFAX:  (301) 619-5034
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11492 base pairs
;      TYPE:  Nucleic acid
;      STRANDEDNESS:  Double
;      TOPOLOGY:  Linear
;    FEATURE:  OTHER INFORMATION:  N at all occurences is = unknown:
;    FEATURE:  OTHER INFORMATION: K at all
;    FEATURE:  occurences is = G or T
US-08-991-840A-1
```

Query Match 89.8%; Score 4492; DB 3; Length 11492;
Best Local Similarity 95.8%; Pred. No. 0;
Matches 4826; Conservative 0; Mismatches 127; Indels 82; Gaps 18;



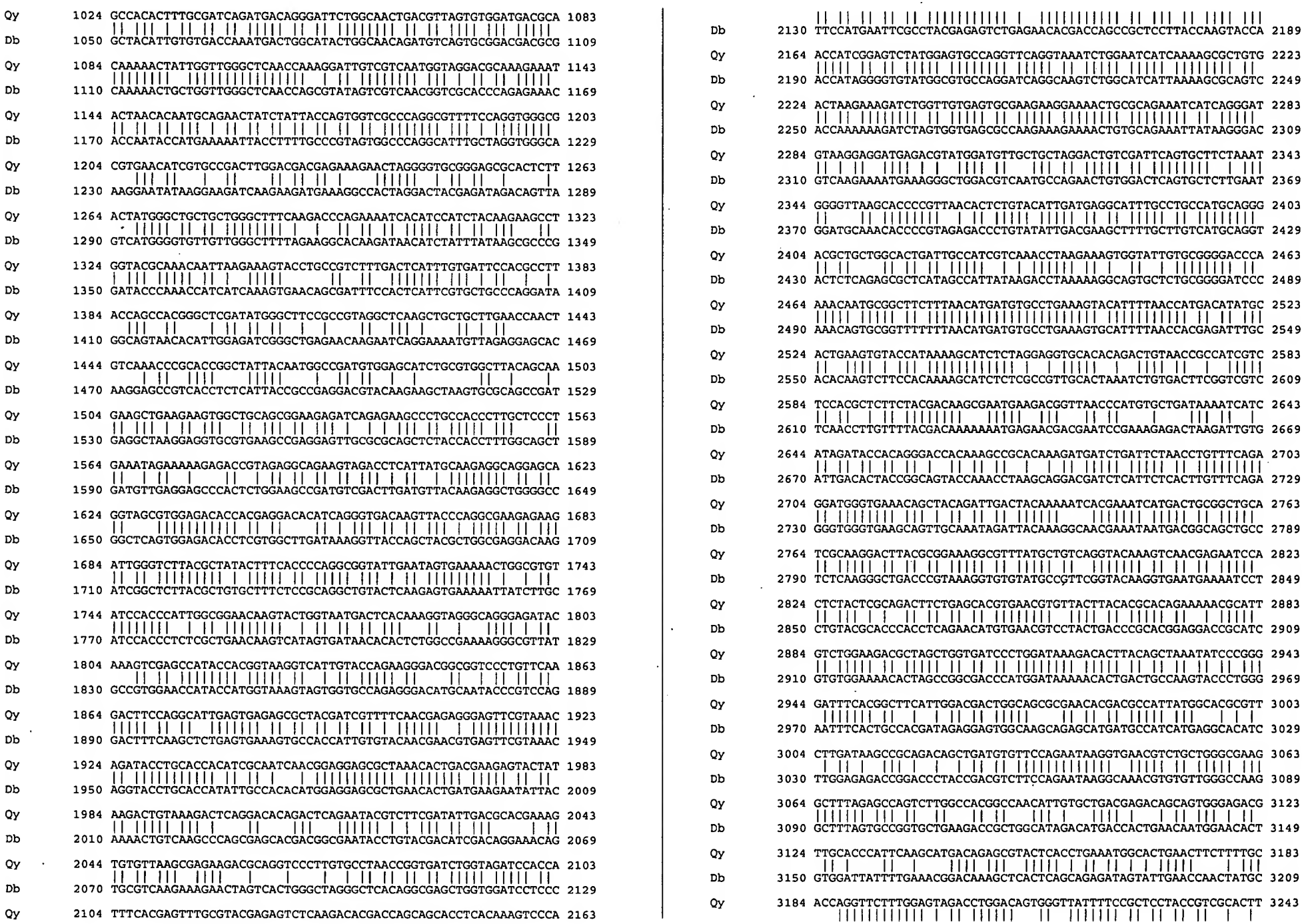



Db 2078 CCTTGTGCCTAACTGGTGATCTGGTAGATCCACCATTTCACGAGTTTGCGTACGAGAGTC 2137
Qy 2129 TCAAGACACGACCAGCAGCACCTCACAAAGTCCCAACCATCGGAGTCTATGGAGTGCCAG 2188
Db 2138 TCAAGACACGACCAGCAGCACCTCACAAAGTCCCAACCATCGGAGTCTATGGAGTGCCAG 2197
Qy 2189 GTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTGACTAAGAAAGATCTGGTTGTGAGTG 2248
Db 2198 GTTCAGGTAAATCTGGGATCATCAAAAGCGCTGTGACTAAGAAAGATCTGGTTGTGAGTG 2257
Qy 2249 CGAAGAAGGAAAACTGCGCAGAAATCATCAGGGATGTAAGGAGGATGAGACGTATGGATG 2308
Db 2258 CGAAGAAGGAAAACTGCGCAGAAATTATCAGGGATGTAAGGAGGATGAGACGTATGGATG 2317
Qy 2309 TTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAATGGGGTTAAGCACCCCGTTAACACTC 2368
Db 2318 TTGCTGCTAGGACTGTTGATTCAGTGCTTCTAAATGGGGTTAAGCACCCCGTTAACACCC 2377
Qy 2369 TGTACATTGATGAGGCATTTGCCTGCCATGCAGGGACGCTGCTGGCACTGATTGCCATCG 2428
Db 2378 TGTACATTGATGAGGCATTTGCCTGCCATGCAGGGACGCTGCTGGCACTGATTGCCATCG 2437
Qy 2429 TCAAACCTAAGAAAGTGGTATTGTGCGGGGACCCAAAACAATGCGGCTTCTTTAACATGA 2488
Db 2438 TCAAACCTAAGAAAGTGGTATTGTGCGGGGACCCAAAACAATGCGGCTTCTTTAACATGA 2497
Qy 2489 TGTGCCTGAAAGTACATTTTAACCATGACATATGCACTGAAGTGTACCATAAAAGCATCT 2548
Db 2498 TGTGCCTGAAAGTACATTTTAACCATGACATATGCACTGAGGTGTACCACAAAAGCATCT 2557
Qy 2549 CTAGGAGGTGCACACAGACTGTAACCGCCATCGTCTCCACGCTCTTCTACGACAAGCGAA 2608
Db 2558 CTAGGAGGTGCACACAGACTGTAACCGCCATTGTCTCCACGCTCTTCTACGACAAGCGAA 2617
Qy 2609 TGAAGACGGTTAACCCATGTGCTGATAAAATCATCATAGATACCACAGGGACCACAAAGC 2668
Db 2618 TGAAGACGGTTAACCCATGTGCTGACAAAATCATCATAGATACCACAGGGACCACAAAGC 2677
Qy 2669 CGCACAAAGATGATCTGATTCTAACCTGTTTCAGAGGATGGGTGAAACAGCTACAGATTG 2728
Db 2678 CGCACAAAGATGATCTGATTCTAACCTGTTTCAGAGGATGGGTGAAACAGCTACAGATTG 2737
Qy 2729 ACTACAAAAATCACGAAATCATGACTGCGGCTGCATCGCAAGGACTTACGCGGAAAGGCG 2788
Db 2738 ACTATAAAAATCATGAAATCATGACTGCGGCTGCATCGCAAGGACTTACGCGGAAAGGCG 2797
Qy 2789 TTTATGCTGTCAGGTACAAAGTCAACGAGAATCCACTCTACTCGCAGACTTCTGAGCACG 2848
Db 2798 TTTATGCTGTCAGGTACAAAGTCAACGAGAATCCACTCTACTCGCAGACTTCTGAGCACG 2857
Qy 2849 TGAACGTGTTACTTACACGCACAGAAAAACGCATTGTCTGGAAGACGCTAGCTGGTGATC 2908
Db 2858 TGAACGTGTTAGTTACACGCACAGAAAAACGCATTGTCTGGAAGACGCTAGCTGGTGACC 2917
Qy 2909 CCTGGATAAAGACACTTACAGCTAAATATCCCGGGGATTTCACGGCTTCATTGGACGACT 2968
Db 2918 CCTGGATAAAGACACTTACAGCTAAATACTCCGGGGATTTCACGGCTTCATTGGACGACT 2977
Qy 2969 GGCAGCGCGAACACGACGCCATTATGGCACGCGTTCTTGATAAGCCGCAGACAGCTGATG 3028
Db 2978 GGCAACGCGAACACGACGCTATTATGGCACGCGTTCTTGATAAGCCGCAGACAGCTGATG 3037
Qy 3029 TGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAGGCTTTAGAGCCAGTCTTGGCCACGG 3088
Db 3038 TGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAGGCTTTAGAGCCAGTCTTGGCCACGG 3097
Qy 3089 CCAACATTGTGCTGACGAGACAGCAGTGGGAGACGTTGCACCCATTCAAGCATGACAGAG 3148
Db 3098 CCAACATTGTGCTGACGAGACAGCAGTGGGAGACGTTGCACCCATTCAAGCATGACAGAG 3157
Qy 3149 CGTACTCACCTGAAATGGCACTGAACTTCTTTTGCACCAGGTTCTTTGGAGTAGACCTGG 3208
Db 3158 CGTACTCACCTGAAATGGCACTGAACTTCTTTTGCACCAGGTTCTTTGGTGTAGACCTGG 3217
Qy 3209 ACAGTGGGTTATTTTCCGCTCCTACCGTCGCACTTACTTACAGGGATCAGCACTGGGATA 3268
Db 3218 ACAGTGGGTTGTTTTCCGCTCCTACCGTCGCACTTACTTACAGGGATCAGCACTGGGACA 3277
Qy 3269 ACTCGCCAGGGAAGAACATGTATGGGCTTAATAGAGAGGTAGCAAAGGAGTTGTCACGGC 3328
Db 3278 ACTCGCCAGGGAAGAACATGTATGGGCTTAATAGAGAGGTAGCAAAGGAGCTGTCACGGC 3337
Qy 3329 GATATCCGTGCATCACAAAAGCGGTTGACACAGGCAGGGTAGCTGATATAAGGAATAATA 3388
Db 3338 GATATCCGTGCATCACAAAAGCGGTTGACACAGGCAGGGTAGCTGATATAAGGAATAATA 3397
Qy 3389 CCATCAAGGACTACTCTCCAACAATTAATGTGGTTCCATTAAATCGCCGGTTGCCCCACT 3448
Db 3398 CCATCAAGGACTACTCTCCAACAATTAATGTGGTTCCATTAAATCGCCGGTTGCCCCACT 3457
Qy 3449 CGTTGATCGTTGACCACAAAGGACAGGGTACAACTGATCACAGCGGATTCCTATCTAAGA 3508
Db 3458 CGCTGATCGTTGACCACAAAGGACAGGGTACAACTGATCACAGTGGATTCCTATCTAAGA 3517
Qy 3509 TGAAGGGCAAATCTGTGTTGGTGATCGGCGATCCTATCAGCATTCCAGGGAAGAAAGTAG 3568
Db 3518 TGAATGGCAAATCTGTGTTGGTGATCGGCGATCCTATCAGCATTCCTGGGAAGAAAGTAG 3577
Qy 3569 AGTCCATGGGTCCATTGCCCACTAATACCATCAGGTGTGATCTCGATTTGGGAATACCTA 3628
Db 3578 AGTCCATGGGTCCATTGCCCACTAATACCATCAGGTGTGATCTCGATTTGGGAATACCTA 3637
Qy 3629 GCCATGTCGGTAAATATGACATTATCTTTGTCAATGTTAGGACCCCGTACAGGAACCATC 3688
Db 3638 GCCATGTCGGTAAATATGACATTATATTTGTCAATGTTAGGACCCCATACAAGAACCATC 3697
Qy 3689 ACTACCAACAGTGCGAGGATCACGCTATCCACCACAGCATGCTAACGTGTAAGGCTGTCC 3748
Db 3698 ACTACCAACAGTGCGAGGATCACGCTATCCACCACAGCATGCTAACGTGTAAGGCTGTCC 3757
Qy 3749 ACCACCTGAACACTGGCGGAACATGTGTGGCTATAGGGTATGGGCTTGCTGATCGCGCAA 3808
Db 3758 ACCACCTGAACACTGGCGGAACATGTGTGGCCATAGGGTATGGACTTGCTGATCGCGCAA 3817
Qy 3809 CCGAGAATATCATCACTGCGGTGGCACGCTCATTTAGGTTTACCCGTGTCTGTCAGCCTA 3868
Db 3818 CCGAGAATATCATCACTGCGGTGGCTCGCTCATTTAGGTTTACCCGTGTCTGTCAGCCTA 3877
Qy 3869 AGAACACTGCCGAAAATACTGAGGTTCTCTTCGTGTTCTTCGGCAAGGACAACGGCAACC 3928
Db 3878 AGAACACTGCCGAAAATACTGAGGTTCTCTTCGTGTTCTTCGGCAAGGACAACGGCAACC 3937
Qy 3929 ACACACATGACCAGGACAGACTCGGTGTAGTGCTTGACAACATCTATCAAGGGTCAACCA 3988
Db 3938 ACACACATGACCAGGACAGACTCGGTGTAGTGCTTGACAACATCTACCAAGGGTCAACCA 3997
Qy 3989 GGTACGAGGCAGGGAGAGCTCCAGCGTACAGAGTGATCAGAGGTGACATTAGCAAGAGCG 4048
Db 3998 GGTACGAGGCAGGGAGAGCTCCAGCGTACAGAGTGATCAGAGGTGACATTAGCAAGAGCG 4057
Qy 4049 CTGACCAAGCTATCGTTAATGCTGCTAATAGCAAAGGTCAACCAGGTTCCGGAGTGTGCG 4108
Db 4058 CTGACCAAGCTATCGTTAATGCTGCTAATAGCAAAGGTCAACCAGGTTCCGGAGTGTGCG 4117
Qy 4109 GTGCACTGTACCGAAAATGGCCGGCTGCTTTTGATAGACAGCCAATAGCTGTCGGGACGG 4168
Db 4118 GTGCACTGTACCGAAAATGGCCGGCTGCTTTTGATAGACAGCCAATAGCTGTCGGGACGG 4177
Qy 4169 CTAGACTTGTGAAGCACGAACCGCTCATCATACATGCTGTAGGACCCAATTTTTCTAAGA 4228
Db 4178 CTAGACTTGTGAAGCACGAACCGCTCATCATACATGCTGTAGGACCGAATTTTTCTAAGA 4237
Qy 4229 TGCCGGAACCGGAGGGCGACCTTAAGCTCGCAGCTGCCTACATGAGCATAGCGTCCATCG 4288
Db 4238 TGCCGGAACGAGAGGGCGACCTTAAGCTCGCAGCTGCCTACATGAGCATAGCGTCCATCG 4297

```
Qy       4289 TCAACGCTGAGCGGATTACAAAAATATCAGTACCGCTACTGTCAACCGGCATCTATTCTG 4348
             |||||||||||||||||||| |||||||||||||||||||||||| ||||||||||||||
Db       4298 TCAACGCTGAGCGGATTACTAAAATATCAGTACCGCTACTGTCAACTGGCATCTATTCTG 4357

Qy       4349 GTGGCAAAGATCGAGTGATGCAATCATTGCATCACCTGTTCACTGCTTTCGACACTACGG 4408
             |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       4358 GTGGCAAAGATCGAGTGATGCAATCATTGCATCACTTGTTCACTGCTTTCGACACTACGG 4417

Qy       4409 ATGCCGATGTCACCATATATTGCTTGGATAAACAATGGGAGACCAGGATAATCGAGGCCA 4468
             |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db       4418 ATGCCGATGTCACCATATATTGCTTGGATAAACAATGGGAGACCAGGATAACCGAGGCCA 4477

Qy       4469 TTCACCGCAAAGAAAGCGTCGAAATTCTG---GATGATGACAAGCCAGTAGACATTGACT 4525
             |||||||||||||||||||||||||| |||    |||||||||||||||||||||||| ||||
Db       4478 TTCACCGCAAAGAAAGCGTCGAAATACTGGATGATGATGACAAGCCAGTAGACATCGACT 4537

Qy       4526 TGGTCAGGGTCCACCCAAACAGCTCTTTGGCAGGCAGACCAGGTTACTCCGTCAATGAGG 4585
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4538 TGGTCAGGGTCCACCCAAACAGCTCTTTGGCAGGCAGACCAGGTTACTCCGTCAATGAGG 4597

Qy       4586 GCAAGTTGTATTCATACCTGGAAGGTACACGATTCCATCAGACCGCCAAGGACATTGCCG 4645
             ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4598 GCAAGCTGTATTCATACCTGGAAGGTACACGATTCCATCAGACCGCCAAGGACATTGCCG 4657

Qy       4646 AAATCCATGCAATGTGGCCCAACAAATCTGAGGCTAATGAGCAGATTTGCTTGTACATCC 4705
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4658 AAATCCATGCAATGTGGCCCAACAAATCTGAGGCTAATGAGCAGATTTGCTTGTACATCC 4717

Qy       4706 TGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGCCCAGTAGAGGAGTCAGAGGCGTCTG 4765
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4718 TGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGCCCAGTAGAGGAGTCAGAGGCGTCTG 4777

Qy       4766 CTCCACCTCACACACTTCCATGCCTGTGTAATTACGCTATGACGGCTGAGCGCGTATACA 4825
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4778 CTCCACCTCACACACTTCCATGCCTGTGTAATTACGCTATGACGGCTGAGCGCGTATACA 4837

Qy       4826 GGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTATGCTCATCATTCCTGTTGCCGAAGT 4885
             |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       4838 GGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTATGCTCATCATTCCCGTTGCCGAAGT 4897

Qy       4886 ACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGCAAACCAGTCCTGTTTTCAGGCGTCG 4945
             |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       4898 ACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGCAAACCAGTCCTGTTCTCAGGCGTCG 4957

Qy       4946 TACCACCGGCTGTACACCCCAGGAAGTACGCGGAAATAATTCTAGAAACGCCACC 5000
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db       4958 TACCACCGGCTGTACACCCCAGGAAGTACGCGGAAATAATTCTAGAAACGCCACC 5012
```

```
RESULT 2
US-09-454-721A-3
; Sequence 3, Application US/09454721A
; Patent No. 6296854
; GENERAL INFORMATION:
;  APPLICANT:   Peter Pushko
;  APPLICANT:  Michael D. Parker
;  APPLICANT:  Jonathan F. Smith
;  APPLICANT:  Burce J. Crise
;  TITLE OF INVENTION:  Live Attenuated Venezuelan Equine Encephalitis Vaccine
;  FILE REFERENCE:  Army 146
; CURRENT APPLICATION NUMBER: US/09/454,721A
;  CURRENT FILING DATE:  1999-12-07
;  PRIOR APPLICATION NUMBER:  US 60/111,330
;  PRIOR FILING DATE:  1998-12-07
;  NUMBER OF SEQ ID NOS:  3
;  SOFTWARE:  Microsoft Word
; SEQ ID NO 3
;   LENGTH: 11459
;   TYPE: DNA
;   ORGANISM: Venezuelan Equine Encephalitis Virus
;   FEATURE:
US-09-454-721A-3
```

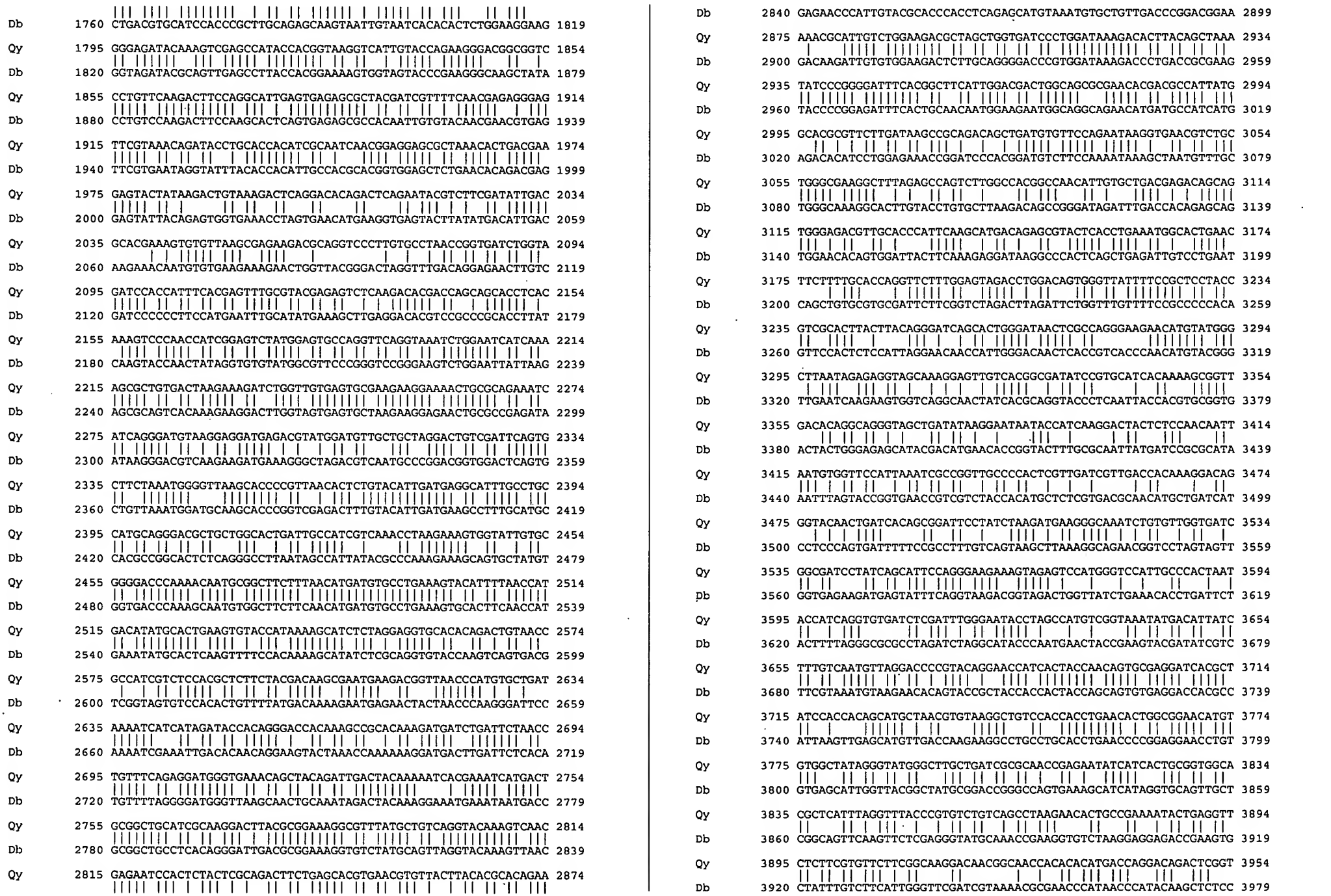
```
  Query Match           42.4%;  Score 2118;  DB 3;  Length 11459;
  Best Local Similarity   64.5%;  Pred. No. 0;
  Matches 3219;  Conservative    0;  Mismatches 1745;  Indels   24;  Gaps    3;

Qy         10 ACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTC 69
              | | | | | |||| ||||| | | |||||||||||| | || |  |||||||| |   ||
Db         30 AATTACCTACCCAAAATGGAGAAAGTTCACGTTGACATCGAGGAAGACAGCCCATTCCTC 89

Qy         70 AAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAAT 129
              |   | || |||||||  || || || |||||| | ||||| | ||||||||||||| |||
Db         90 AGAGCTTTGCAGCGGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCACTGATAAT 149

Qy        130 GACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTC 189
              |||||||| |||||||||||||||||||| |||| |||| || || || |  || ||
Db        150 GACCATGCTAATGCCAGAGCGTTTTCGCATCTGGCTTCAAAACTGATCGAAACGGAGGTG 209

Qy        190 GACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAAT 249
              ||||    | |    ||| | ||||||||||||||||||| | |         |||| ||
Db        210 GACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCCGCCCGCAGAATGTATTCTAAG 269

Qy        250 CACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGG 309
              |||   |||||||||||||||| || |||| | | || |||||  |||||||||  |  |  |
Db        270 CACAAGTATCATTGTATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGATTGTATAAG 329

Qy        310 TATGCAGAAAGACTTAAGAAAAGT------GACATTACCGACAAGAACATAGCCTCTAAG 363
              ||||||    |  || |||||||          || || || || ||| |  | | |   ||
Db        330 TATGCAACTAAGCTGAAGAAAAACTGTAAGGAAATAACTGATAAGGAATTGGACAAGAAA 389

Qy        364 GCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATG 423
                |  || ||    |  ||||||      || |||   || |||    || ||||  |
Db        390 ATGAAGGAGCTCGCCGCCGTCATGAGCGACCCTGACCTGGAAACTGAGACTATGTGCCTC 449

Qy        424 CACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCA 483
              |||  ||||  ||||| | |||   ||    || || || ||||| ||||| |||||
Db        450 CACGACGACGAGTCGTGTCGCTACGAAGGGCAAGTCGCTGTTTACCAGGATGTATACGCG 509

Qy        484 GTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGG 543
              ||  | | |||||||    |||| ||||| ||   ||| ||||||||       ||||||
Db        510 GTTGACGGACCGACAAGTCTCTATCACCAAGCCAATAAGGGAGTTAGAGTCGCCTACTGG 569

Qy        544 ATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTAC 603
              |||||||||||||| ||||||||||||   || ||| |||| ||  | || ||  | |||
Db        570 ATAGGCTTTGACACCACCCCTTTTATGTTTAAGAACTTGGCTGGAGCATATCCATCATAC 629

Qy        604 AACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCA 663
                || |||||||| |||||  | || ||    || |||||||| |||||  | | |||
Db        630 TCTACCAACTGGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCCTATGCAGCTCT 689

Qy        664 GATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCT 723
              ||  ||  |||| |    | | ||    | || || ||||| ||||||     |  ||||
Db        690 GACGTTATGGAGCGGTCACGTAGAGGGATGTCCATTCTTAGAAAGAAGTATTTGAAACCA 749

Qy        724 ACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTA 783
              | || ||  |  ||||||| ||||| || || ||||||   ||  | ||     |  |
Db        750 TCCAACAATGTTCTATTCTCTGTTGGCTCGACCATCTACCACGAGAAGAGGGACTTACTG 809

Qy        784 CGTAGCTGGCATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGA 843
              | |||||||| || ||    || || |||||     || |||   || | |||| || |
Db        810 AGGAGCTGGCACCTGCCGTCTGTATTTCACTTACGTGGCAAGCAAAATTACACATGTCGG 869

Qy        844 TGTGGGACCATTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGA 903
              |||| ||| || || || || || ||||||||| | ||||  ||| | ||||| |||||
Db        870 TGTGAGACTATAGTTAGTTGCGACGGGTACGTCGTTAAAAGAATAGCTATCAGTCCAGGC 929

Qy        904 CTATACGGTAAAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGC 963
              || ||. || ||    |      || ||  | || ||||| |||||||| |||||| | |||
Db        930 CTGTATGGGAAGCCTTCAGGCTATGCTGCTACGATGCACCGCGAGGGATTCTTGTGCTGC 989

Qy        964 AAAGTCACAGATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCA 1023
              ||||| ||||| ||  ||  ||| |||||||| |||||| | ||||| |||||||| |||
Db        990 AAAGTGACAGACACATTGAACGGGGAGAGGGTCTCTTTTCCCGTGTGCACGTATGTGCCA 1049
```



```
Db       3210 GTGAGGTTCTTTGGACTCGATCTGGACTCCGGTCTATTTTCTGCACCCACTGTTCCGTTA 3269
Qy       3244 ACTTACAGGGATCAGCACTGGGATAACTCGCCAGGGAAGAACATGTATGGGCTTAATAGA 3303
Db       3270 TCCATTAGGAATAATCACTGGGATAACTCCCCGTCGCCTAACATGTACGGGCTGAATAAA 3329
Qy       3304 GAGGTAGCAAAGGAGTTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACAGGC 3363
Db       3330 GAAGTGGTCCGTCAGCTCTCTCGCAGGTACCCACAACTGCCTCGGGCAGTTGCCACTGGA 3389
Qy       3364 AGGGTAGCTGATATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTGGTT 3423
Db       3390 AGAGTCTATGACATGAACACTGGTACACTGCGCAATTATGATCCGCGCATAAACCTAGTA 3449
Qy       3424 CCATTAAATCGCCGGTTGCCCCACTCGTTGATCGTTGACCACAAAGGACAGGGTACAACT 3483
Db       3450 CCTGTAAACAGAAGACTGCCTCATGCTTTAGTCCTCCACCATAATGAACACCCACAGAGT 3509
Qy       3484 GATCACAGCGGATTCCTATCTAAGATGAAGGGCAAATCTGTGTTGGTGATCGGCGATCCT 3543
Db       3510 GACTTTTCTTCATTCGTCAGCAAATTGAAGGGCAGAACTGTCCTGGTGGTCGGGGAAAAG 3569
Qy       3544 ATCAGCATTCCAGGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATACCATCAGG 3603
Db       3570 TTGTCCGTCCCAGGCAAAATGGTTGACTGGTTGTCAGACCGGCCTGAGGCTACCTTCAGA 3629
Qy       3604 TGTGATCTCGATTTGGGAATACCTAGCCATGTCGGTAAATATGACATTATCTTTGTCAAT 3663
Db       3630 GCTCGGCTGGATTTAGGCATCCCAGGTGATGTGCCCAAATATGACATAATATTTGTTAAT 3689
Qy       3664 GTTAGGACCCCGTACAGGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCACCAC 3723
Db       3690 GTGAGGACCCCATATAAATACCATCACTATCAGCAGTGTGAAGACCATGCCATTAAGCTT 3749
Qy       3724 AGCATGCTAACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCTATA 3783
Db       3750 AGCATGTTGACCAAGAAAGCTTGTCTGCATCTGAATCCCGGCGGAACCTGTGTCAGCATA 3809
Qy       3784 GGGTATGGGCTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCACGCTCATTT 3843
Db       3810 GGTTATGGTTACGCTGACAGGGCCAGCGAAAGCATCATTGGTGCTATAGCGCGGCAGTTC 3869
Qy       3844 AGGTTTACCCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTCGTG 3903
Db       3870 AAGTTTTCCCGGGTATGCAAACCGAAATCCTCACTTGAAGAGACGGAAGTTCTGTTTGTA 3929
Qy       3904 TTCTTCGGCAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTGCTT 3963
Db       3930 TTCATTGGGTACGATCGCAAGGCCCGTACGCACAATCCTTACAAGCTTTCATCAACCTTG 3989
Qy       3964 GACAACATCTATCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGTACAGAGTG 4023
Db       3990 ACCAACATTTATACAGGTTCCAGACTCCACGAAGCCGGATGTGCACCCTCATATCATGTG 4049
Qy       4024 ATCAGAGGTGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGCAAA 4083
Db       4050 GTGCGAGGGGATATTGCCACGGCCACCGAAGGAGTGATTATAAATGCTGCTAACAGCAAA 4109
Qy       4084 GGTCAACCAGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTTGAT 4143
Db       4110 GGACAACCTGGCGGAGGGGTGTGCGGAGCGCTGTATAAGAAGTTCCCGGAAAGCTTCGAT 4169
Qy       4144 AGACAGCCAATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACC---GCTCATCATA 4200
Db       4170 TTACAGCCGATCGAAGTAGGAAAAGCGCGACTGGTCAAAGGTGCAGCTAAACATATCATT 4229
Qy       4201 CATGCTGTAGGACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTAAGCTCGCA 4260
Db       4230 CATGCCGTAGGACCAAACTTCAACAAAGTTTCGGAGGTTGAAGGTGACAAACAGTTGGCA 4289
Qy       4261 GCTGCCTACATGAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTA 4320
Db       4290 GAGGCTTATGAGTCCATCGCTAAGATTGTCAACGATAACAATTACAAGTCAGTAGCGATT 4349

Qy       4321 CCGCTACTGTCAACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCAT 4380
Db       4350 CCACTGTTGTCCACCGGCATCTTTTCCGGAAACAAAGATCGACTAACCCAATCATTGAAC 4409
Qy       4381 CACCTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAA 4440
Db       4410 CATTTGCTGACAGCTTTAGACACCACTGATGCAGATGTAGCCATATACTGCAGGGACAAG 4469
Qy       4441 CAATGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATTCTGGA- 4499
Db       4470 AAATGGGAAATGACTCTCAAGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGGAGATATGC 4529
Qy       4500 --------------TGATGACAAGCCAGTAGACATTGACTTGGTCAGGGTCCACCCAAAC 4545
Db       4530 ATATCCGACGACTCTTCAGTGACAGAACCTGATGCAGAGCTGGTGAGGGTGCATCCGAAG 4589
Qy       4546 AGCTCTTTGGCAGGCAGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTG 4605
Db       4590 AGTTCTTTGGCTGGAAGGAAGGGCTACAGCACAAGCGATGGCAAAACTTTCTCATATTTG 4649
Qy       4606 GAAGGTACACGATTCCATCAGACCGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCC 4665
Db       4650 GAAGGGACCAAGTTTCACCAGGCGGCCAAGGATATAGCAGAAATTAATGCCATGTGGCCC 4709
Qy       4666 AACAAATCTGAGGCTAATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGC 4725
Db       4710 GTTGCAACGGAGGCCAATGAGCAGGTATGCATGTATATCCTCGGAGAAAGCATGAGCAGT 4769
Qy       4726 ATCCGCTCCAAATGCCCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCA 4785
Db       4770 ATTAGGTCGAAATGCCCCGTCGAAGAGTCGGAAGCCTCCACACCACCTAGCACGCTGCCT 4829
Qy       4786 TGCCTGTGTAATTACGCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAA 4845
Db       4830 TGCTTGTGCATCCATGCCATGACTCCAGAAAGAGTACAGCGCCTAAAAGCCTCACGTCCA 4889
Qy       4846 GAACAGTTCGCCGTATGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAG 4905
Db       4890 GAACAAATTACTGTGTGCTCATCCTTTCCATTGCCGAAGTATAGAATCACTGGTGTGCAG 4949
Qy       4906 AAGCTACAGTGCAGCAAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCC 4965
Db       4950 AAGATCCAATGCTCCCAGCCTATATTGTTCTCACCGAAAGTGCCTGCGTATATTCATCCA 5009
Qy       4966 AGGAAGTA 4973
Db       5010 AGGAAGTA 5017
```

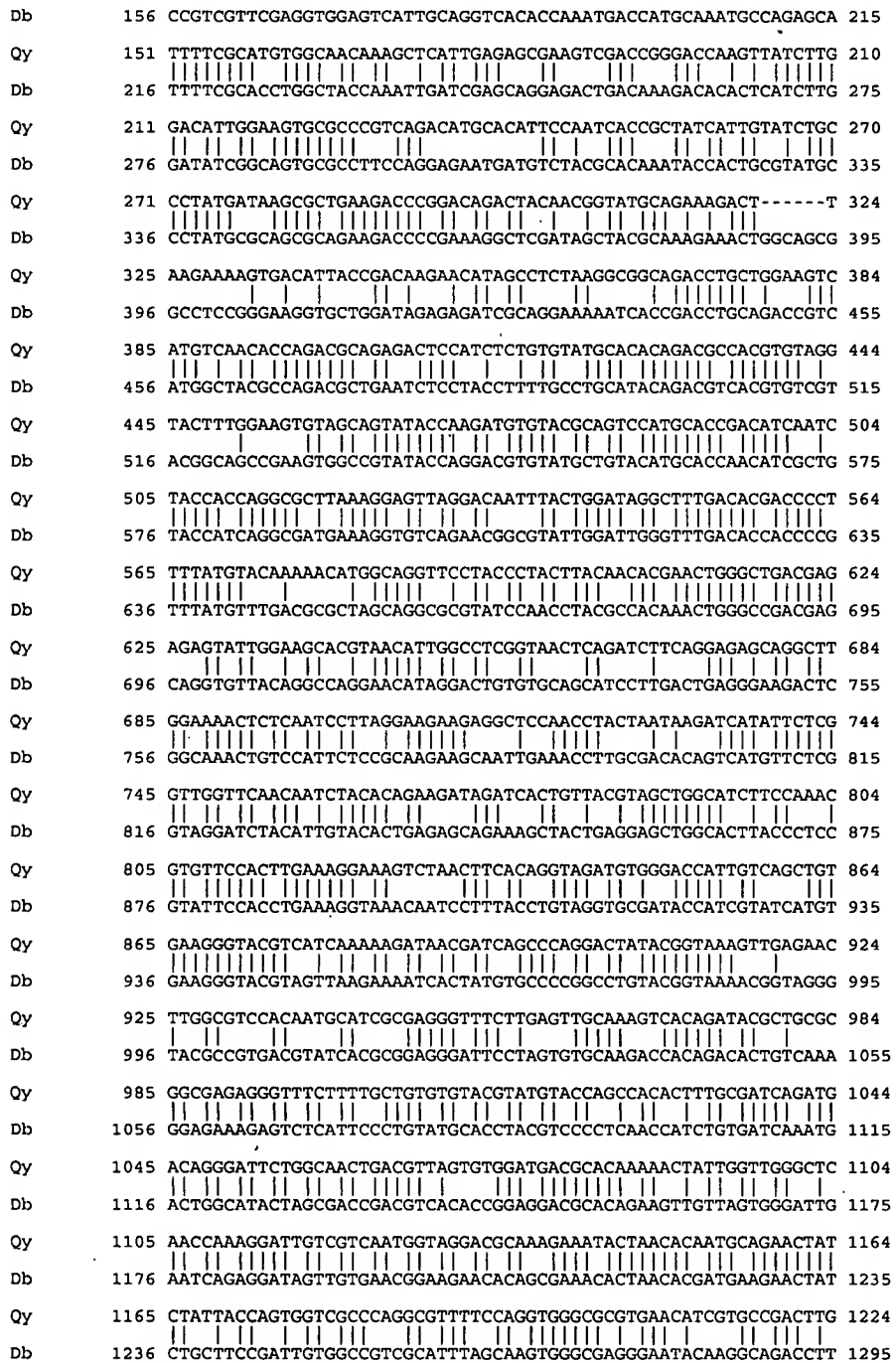
RESULT 3
US-08-991-840A-2
; Sequence 2, Application US/08991840A
; Patent No. 6261570
; GENERAL INFORMATION:
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Bruce Crise
; APPLICANT: Mark Steve Oberste
; APPLICANT: Shannon Schmura
; TITLE OF INVENTION: Live Attenuated Virus Vaccines for Eastern Equine Encephalitis
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Charles H. Harris
; STREET: USA MRMC - MRMC-JA
; CITY: FORT DETRICK, FREDERICK
; STATE: MARYLAND
; COUNTRY: USA
; ZIP: 21702-5012
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.5

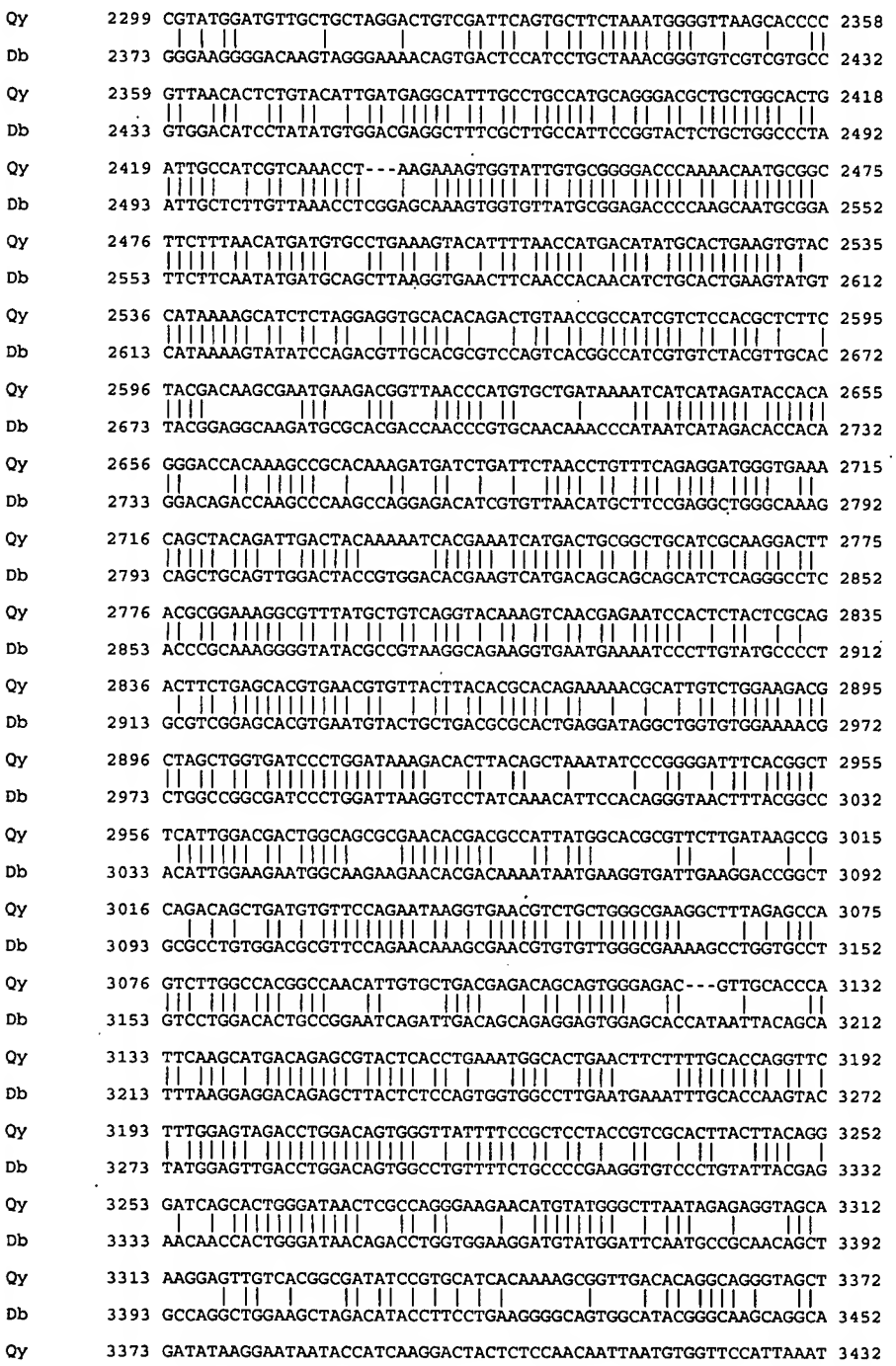
```
;      SOFTWARE:  Microsoft Word 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/991,840A
;      FILING DATE:  December 16, 1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  Provisional Application 60/047162,
;      FILING DATE:  May 20, 1997
;      APPLICATION NUMBER:  Provisional Application 60/053,652
;      FILING DATE:  July 24, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles H. Harris
;      REGISTRATION NUMBER:  34,616
;      REFERENCE/DOCKET NUMBER:  003/058/SAP RIID 96-01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 619-2065
;      TELEFAX:  (301) 619-5034
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11464 base pairs
;      TYPE:  Nucleic acid
;      STRANDEDNESS:  Double
;      TOPOLOGY:  Unknown
US-08-991-840A-2

  Query Match             41.5%;  Score 2074.8;  DB 3;  Length 11464;
  Best Local Similarity   64.1%;  Pred. No. 0;
  Matches 3202;  Conservative    0;  Mismatches 1772;  Indels   24;  Gaps    4;

Qy          1 ACCCTACAAACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGC 60
              | || | |      | | | |||| ||||| | | ||||||||||| | || |  || ||
Db         20 AGCCCAAACCAATAACTACCCAAAATGGAGAAAGTTCACGTTGACATCGAGGAAGATAGT 79

Qy         61 CCGTATGTCAAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTC 120
              ||  |   |||   | ||||| ||||  || || || |||||| | ||||| | |||||||
Db         80 CCCTTCCTCAGAGCATTACAACGGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTC 139

Qy        121 ACTGACAATGACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAG 180
              || || |||||||||||| || ||||||||||||||||||| ||||| | ||  | || |||
Db        140 ACAGATAATGACCATGCTAACGCCAGAGCGTTTTCGCATTTGGCATCGAAATTGATCGAG 199

Qy        181 AGCGAAGTCGACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCA 240
              |  || || || |    | |    ||| | ||||||||||||||||| | | |
Db        200 ACGGAGGTGGAACCATCCGATACGATCCTAGACATTGGAAGTGCGCCTGCCCGCAGAATG 259

Qy        241 CATTCCAATCACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGA 300
              ||||||| ||     || ||||| ||||| || |||| | | || ||||| |||||||||
Db        260 TATTCCAAGCATAAGTACCATTGCATCTGTCCGATGAAATGTGCAGAAGATCCGGACAGA 319

Qy        301 CTACAACGGTATGCAGAAAGACTTAAGAAAAGT------GACATTACCGACAAGAACATA 354
              ||       |||||||  |  || ||||| |         || ||||| || ||| |  |
Db        320 CTGTTTAAGTATGCAGCCAAGCTGAAGAAGAACTGTAAAGAGATTACAGATAAGGAACTG 379

Qy        355 GCCTCTAAGGCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCT 414
              | |   |||  |   || ||   ||||||||||       || ||    || |||  | |
Db        380 GACAAGAAGATGAAGGAGCTTGCGGAAGTCATGAGCGACCCTGATCTCGAAACTGAAACG 439

Qy        415 CTGTGTATGCACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGAT 474
              | ||  | |||    || | || ||| | |      ||    || ||||| || || |||
Db        440 ATTTGCCTTCACGACGATGAAACCTGTCGATTTGAGGGTCAAGTCGCAGTGTATCAGGAT 499

Qy        475 GTGTACGCAGTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACA 534
              |||||||| ||  | | ||||||     ||||| |||||     ||||| || ||
Db        500 GTGTACGCGGTTGACGGACCGACGAGCCTTTACCATCAGGCCAACAAAGGGGTCAGAGTC 559

Qy        535 ATTTACTGGATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTAC 594
                || |||||||| || ||||| ||||||||||||  || ||| |||| ||  | ||
Db        560 GCCTATTGGATAGGATTCGACACTACCCCTTTTATGTTTAAGAACCTGGCTGGAGCATAT 619

Qy        595 CCTACTTACAACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTC 654
              ||  | ||    || |||||||| |||||||  || ||    || ||||| || ||| |
Db        620 CCCTCCTATTCGACCAACTGGGCCGACGAGACCGTGTTAACGGCTCGTAATATAGGCTTG 679

Qy        655 GGTAACTCAGATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGG 714
              | | ||| ||| |   |||| |     | | ||     | || ||||| || |||||
Db        680 TGCAGCTCCGATGTGATGGAGCGGTCACGTAGAGGGATGTCCATCCTCAGAAAGAAATTT 739

Qy        715 CTCCAACCTACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGA 774
              |  ||||  | |||||  || | |||||| || || || || ||||||   ||  |  ||
Db        740 TTAAAACCATCCAATAACGTCTTGTTCTCTGTAGGATCTACCATCTACCACGAGAAGCGA 799

Qy        775 TCACTGTTACGTAGCTGGCATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTC 834
              |  || | || ||||| || ||   ||| || ||| |   || |||   || | |
Db        800 GACTTACTAAGGAGTTGGCACCTACCGTCCGTTTTTCACCTACGTGGTAAGCAGAATTAC 859

Qy        835 ACAGGTAGATGTGGGACCATTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATC 894
              ||  || | |||| ||| || || || || || ||||||||| |||||| |||| | ||
Db        860 ACTTGTCGGTGTGAGACTATAGTTAGTTGCGACGGGTACGTCGTCAAAAGGATAGCTATT 919

Qy        895 AGCCCAGGACTATACGGTAAAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTC 954
              || ||||| || ||||| |||      |  ||  ||  |||| ||||||||||||||| |||
Db        920 AGTCCAGGTCTGTACGGGAAACCGTCGGGCTACGCTGCCACGATGCATCGCGAGGGATTC 979

Qy        955 TTGAGTTGCAAAGTCACAGATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACG 1014
              ||| | ||||| || || || |||||   ||| |||||||| |||||  | || || |||
Db        980 TTGTGCTGCAAGGTGACGGACACGCTTGACGGGGAGAGGGTCTCTTTCCCCGTATGCACG 1039

Qy       1015 TATGTACCAGCCACACTTTGCGATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTG 1074
              || || |||||||||| | ||||| || |||||||| |||||||||| || || ||||
Db       1040 TACGTGCCAGCCACATTGTGCGACCAAATGACAGGCATTCTGGCAACAGATGTCAGTGCA 1099

Qy       1075 GATGACGCACAAAAACTATTGGTTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACG 1134
              || ||||| || ||| |  ||||||||||||||||  | || |||||  |||||  | ||
Db       1100 GACGACGCGCAGAAATTGCTGGTTGGGCTCAACCAGCGAATAGTCGTAAATGGCCGCACT 1159

Qy       1135 CAAAGAAATACTAACACAATGCAGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCC 1194
              ||||| |||||||||||||||| |||||||| |||||||| || || || || || ||| ||
Db       1160 CAAAGGAATACTAACACAATGAAGAACTATTTATTACCCGTCGTGGCACAAGCATTTGCC 1219

Qy       1195 AGGTGGGCGCGTGAACATCGTGCCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAG 1254
              || |||||    ||  |    |  ||    || || || |  | | ||| |  ||||
Db       1220 AGATGGGCTAAAGAGTACACAGAAGATCAAGAAGATGAAAGACCATTGGGGCTTAGGGAC 1279

Qy       1255 CGCACTCTTACTATGGGCTGCTGCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTAC 1314
              |||    |    |||| || || ||||| |||| ||  || || || ||||| | |||
Db       1280 CGCCAGTTGGTAATGGGGTGTTGTTGGGCGTTCAGGAAACACAAGATAACATCAGTGTAC 1339

Qy       1315 AAGAAGCCTGGTACGCAAACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATT 1374
              ||   || |  || |||| || |  || |||     ||  ||  |||| || ||| |
Db       1340 AAACGACCAGACACCCAAACGATCATCAAGGTAAACAGCGATTTCCACTCTTTCGTGCTG 1399

Qy       1375 CCACGCCTTACCAGCCACGGGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTT 1434
              ||  |  ||  ||| ||  | || || |  |    ||| |||    || ||
Db       1400 CCCAGAATTGGAAGCAACACCTTGGAGATTGGGCTGAGGACCAGGATCAGAAAACTACTG 1459

Qy       1435 GAACCAACTGTCAAACCCGCACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGC 1494
              ||  | |||  |    ||||  |||||  | |||  |  || |      |
Db       1460 GAGGAACCTGTGGACAGACCACCACTGATTACCGCCGACGACATACAGGAAGCCAAGAAC 1519

Qy       1495 TTACAGCAAGAAGCTGAAGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCC 1554
              |  | || ||| | |||||     |  || |||||| |||| | |||  | |||||
Db       1520 GCGGCGGATGAGGCTAAGGAAGTCAAGGAGGCCGAAGAGCTCAGGGCAGCATTACCACCG 1579

Qy       1555 TTGCTCCCTGAAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAG 1614
              ||    | ||  |||| ||    | | ||||| || ||| | ||  |||| |||
Db       1580 CTGTCTGCCGATGTAGAGGAACCTGCACTGGAGGCGGACGTTGACTTAATGCTGCAGGAG 1639

Qy       1615 GCAGGAGCAGGTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGC 1674
              || ||||||||    || || || ||  | || | ||||  || || |||||  ||||
Db       1640 GCGGGAGCAGGATCTGTCGAAACCCCCAGGGGGCTCATCAAAGTCACCAGTTATGCAGGA 1699

Qy       1675 GAAGAGAAGATTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAA 1734
              ||||| || ||||| ||||| ||  | || || || || || || ||| ||   |||||||||
Db       1700 GAAGACAAAATTGGCTCTTATGCGGTGCTCTCCCCACAAGCAGTACTGCGAAGTGAAAAA 1759

Qy       1735 CTGGCGTGTATCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCA 1794
```



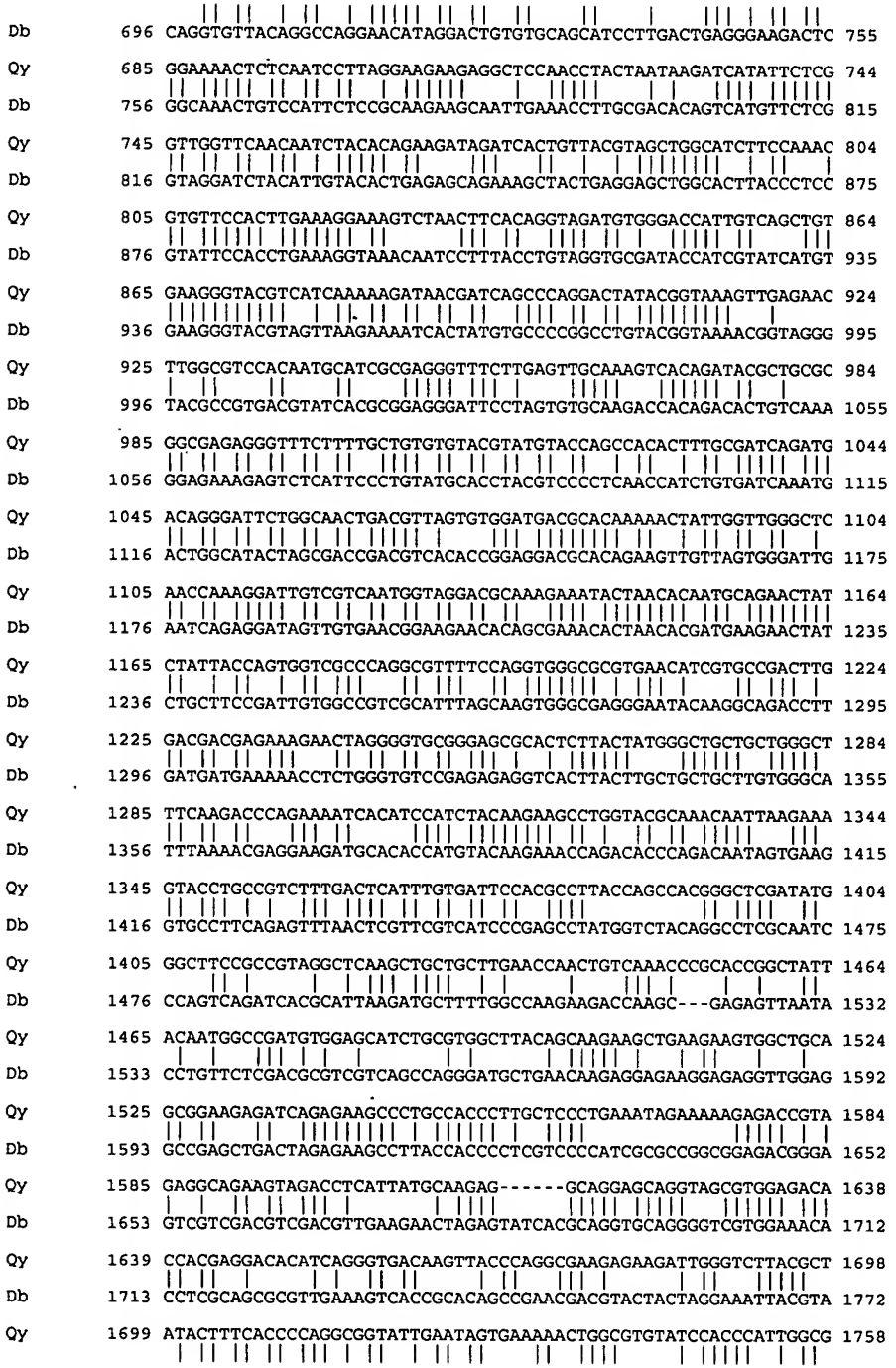
```
Qy       3955 GTAGTGCTTGACAACATCTATCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCG 4014
Db       3980 TCCACCCTGACCAATATTTACACCGGCTCGAGGCTCCATGAAGCTGGCTGCGCACCTTCG 4039

Qy       4015 TACAGAGTGATCAGAGGTGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCT 4074
Db       4040 TATCATGTAGTGCGCGGGGATATAGCCACTGCCACGGAAGGAGTAATCGTTAATGCTGCC 4099

Qy       4075 AATAGCAAAGGTCAACCAGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCT 4134
Db       4100 AACAGCAAGGGCCAGCCAGGCAGTGGAGTGTGCGGAGCTCTGTACCGGAAGTACCCCGAA 4159

Qy       4135 GCTTTTGATAGACAGCCAATAGCTGTCGGGACGGCTAGACTTGTGAA---GCACGAACCG 4191
Db       4160 AGCTTCGATTTACAACCAATAGAAGTGGGGAAAGCTAGATTGGTCAAAGGTAACTCAAAA 4219

Qy       4192 CTCATCATACATGCTGTAGGACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTT 4251
Db       4220 CATCTCATTCATGCAGTGGGGCCGAATTTTAACAAAGTGTCTGAAGTGGAAGGTGACAAA 4279

Qy       4252 AAGCTCGCAGCTGCCTACATGAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAA 4311
Db       4280 CAGCTGGCAGAAGCGTATGAATCTATAGCCAGGATTATTAATGACAACAATTATAGATCT 4339

Qy       4312 ATATCAGTACCGCTACTGTCAACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAA 4371
Db       4340 GTGGCTATTCCGCTTCTGTCCACTGGAATATTTGCCGGAAACAAGGATAGGTTAATGCAA 4399

Qy       4372 TCATTGCATCACCTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGC 4431
Db       4400 TCCTTAAACCATCTGTTAACGGCATTGGACACAACAGACGCAGATGTGGCCATATACTGC 4459

Qy       4432 TTGGATAAACAATGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGA- 4490
Db       4460 AGAGACAAGAAATGGGAAGTGACGTTGAAAGAGGTCGTAGCCAGGAGAGAGGCGGTAGAG 4519

Qy       4491 --------AATTCTGGATGATGACAAGCCAGTAGACATTGACT------TGGTCAGGGTC 4536
Db       4520 GAGATATGTATCTCCGAAGATTCCTCCGTAGCAGAGCCGGATGCAGAGCTGGTTAGAGTT 4579

Qy       4537 CACCCAAACAGCTCTTTGGCAGGCAGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTAT 4596
Db       4580 CACCCTAAGAGCTCTTTGGCTGGAAGGAAAGGTTACAGCACTAGCGATGGGAAGACATTC 4639

Qy       4597 TCATACCTGGAAGGTACACGATTCCATCAGACCGCCAAGGACATTGCCGAAATCCATGCA 4656
Db       4640 TCATATCTTGAAGGAACCAAATTTCATCAGGCGGCGAAGGACATGGCAGAAATTAACGCT 4699

Qy       4657 ATGTGGCCCAACAAATCTGAGGCTAATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGT 4716
Db       4700 ATGTGGCCTGCCGCTACAGAGGCTAATGAGCAGGTGTGCTTATACATTCTGGGTGAAAGT 4759

Qy       4717 ATGTCCAGCATCCGCTCCAAATGCCCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCAC 4776
Db       4760 ATGAGCAGTATAAGATCCAAATGCCCCGTTGAGGAGTCAGAGGCATCCACCCCACCAAGT 4819

Qy       4777 ACACTTCCATGCCTGTGTAATTACGCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCT 4836
Db       4820 ACATTGCCTTGCTTGTGCATCCACGCTATGACCCCGGAACGGGTTCAGCGTTTGAAAGCC 4879

Qy       4837 GCGAAGAAAGAACAGTTCGCCGTATGCTCATCATTCCTGTTGCCGAAGTACAGGATCACA 4896
Db       4880 TCCCGCCCCGAACAAATTACAGTTTGTTCTTCCTTCCCATTGCCGAAGTACAGAATAACA 4939

Qy       4897 GGCGTGCAGAAGCTACAGTGCAGCAAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCT 4956
Db       4940 GGAGTGCAGAAGATTCAATGTTCGCATCCTATACTTTTCTCTCCTAAAGTACCTGAGTAC 4999

Qy       4957 GTACACCCCAGGAAGTAC 4974
Db       5000 ATACACCCTAGAAAGTAC 5017
```

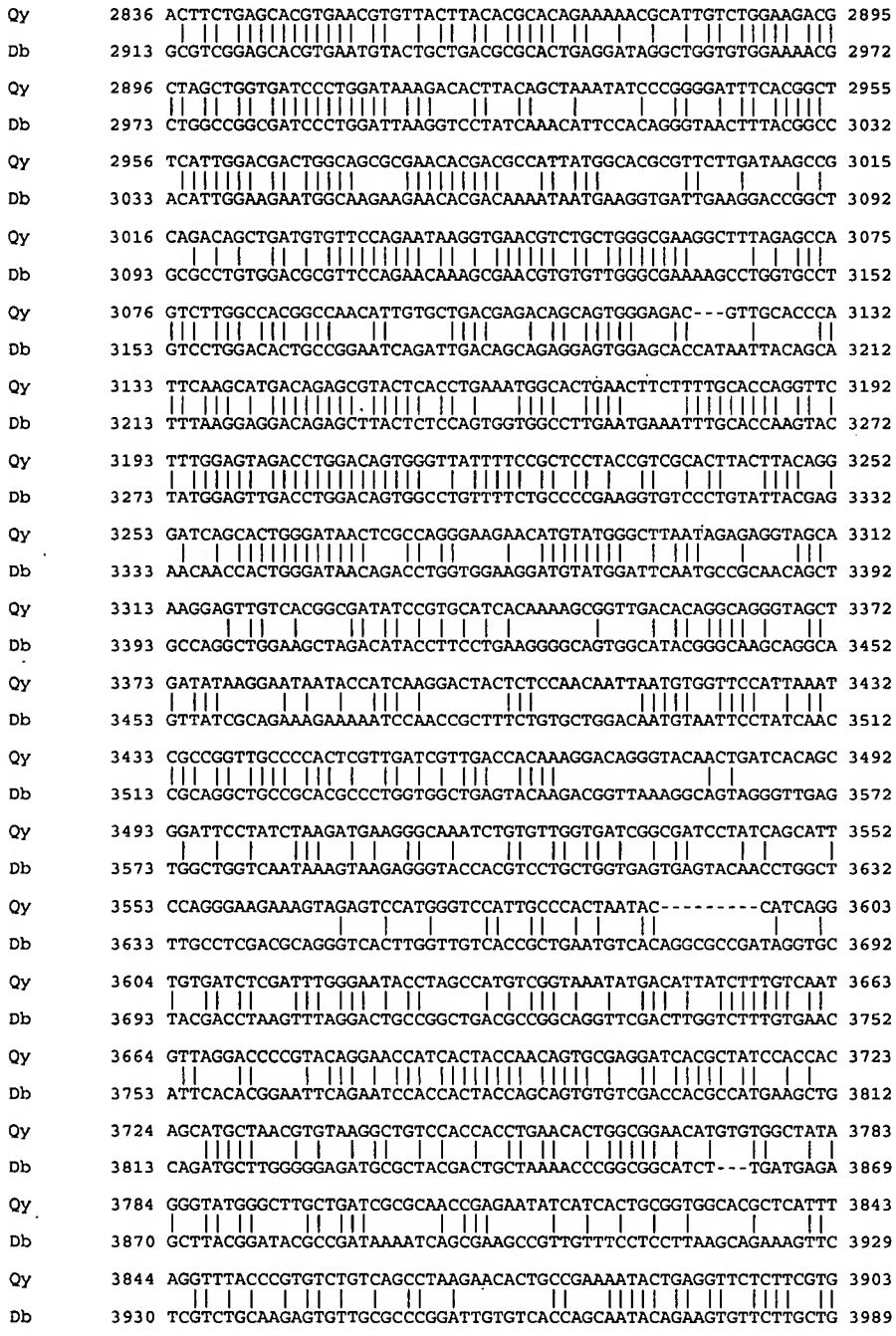
```
RESULT 4
US-07-920-281C-1
; Sequence 1, Application US/07920281C
; Patent No. 5739026
;  GENERAL INFORMATION:
;    APPLICANT:  Garoff, Henrik
;    APPLICANT:  Liljestrom, Peter
;    TITLE OF INVENTION:  DNA Expression Systems Based on
;    TITLE OF INVENTION:  Alphaviruses
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-0747
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/920,281C
;      FILING DATE:  13-AUG-1992
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  828-103P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-241-1300
;      TELEFAX:  703-241-2848
;      TELEX:  248345
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11517 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  RNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Semliki Forest Virus
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  1..11517
;      OTHER INFORMATION:  /label= genome
;      OTHER INFORMATION:  /note= "Semliki Forest Virus complete nucleotide
;      OTHER INFORMATION:  sequence, presented as a cloned DNA sequence; see
;      OTHER INFORMATION:  Figure 5."
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  87..7379
;      OTHER INFORMATION:  /product= "SFV polyprotein"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  7421..11179
;      OTHER INFORMATION:  /product= "SFV polyprotein"
US-07-920-281C-1

  Query Match           31.3%;  Score 1564;  DB 1;  Length 11517;
  Best Local Similarity  58.3%;  Pred. No. 0;
  Matches 2904;  Conservative    0;  Mismatches 2040;  Indels   36;  Gaps    8;

Qy         31 AGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGTCGTTACAGCGGACGTTT 90
Db         96 AAAGTGCATGTTGATATTGAGGCTGACAGCCCATTCATCAAGTCTTTGCAGAAGGCATTT 155

Qy         91 CCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACCATGCCAATGCCAGAGCG 150
```

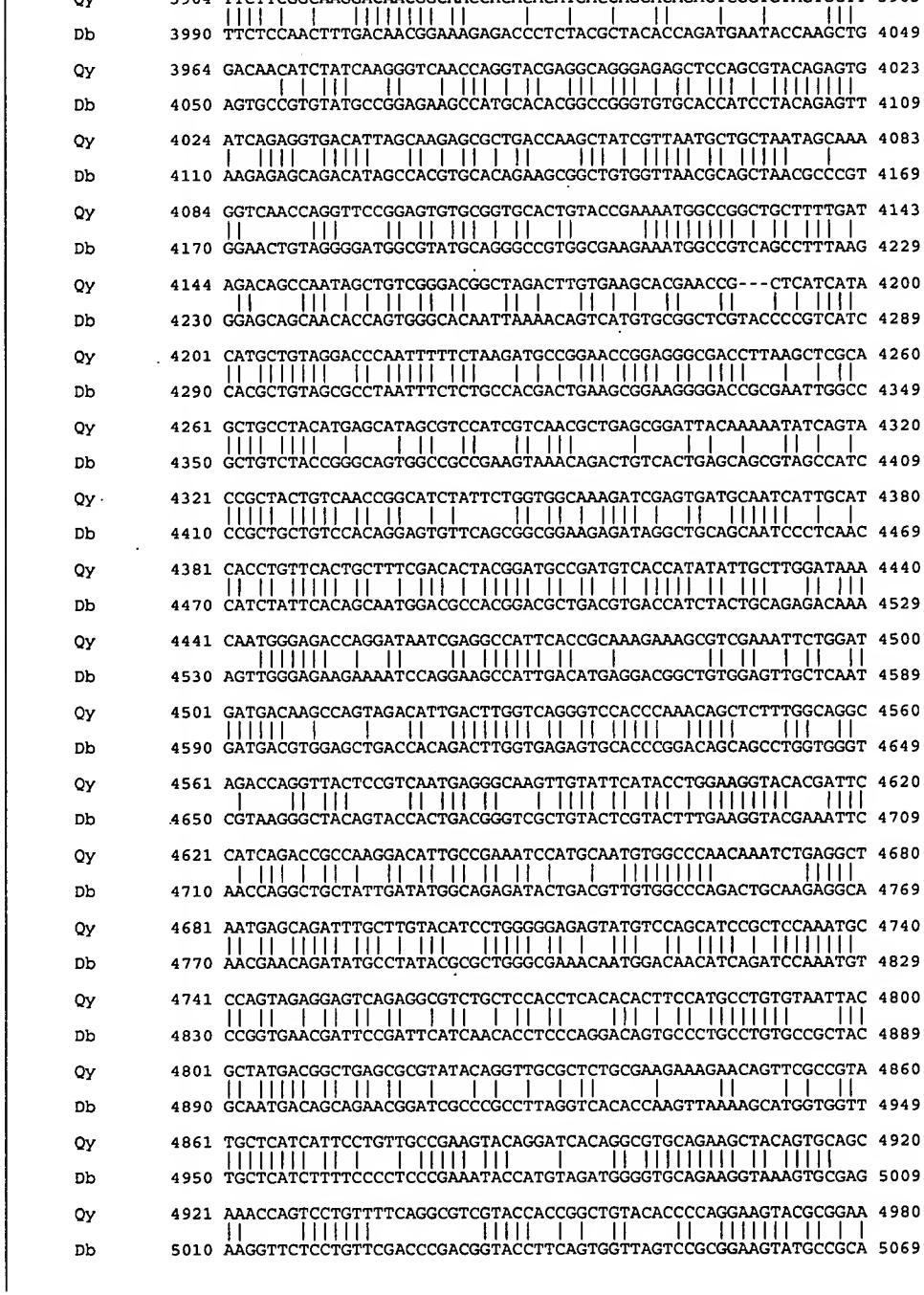

Db 156 CCGTCGTTCGAGGTGGAGTCATTGCAGGTCACACCAAATGACCATGCAAATGCCAGAGCA 215
Qy 151 TTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACCGGGACCAAGTTATCTTG 210
Db 216 TTTTCGCACCTGGCTACCAAATTGATCGAGCAGGAGACTGACAAAGACACACTCATCTTG 275
Qy 211 GACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACCGCTATCATTGTATCTGC 270
Db 276 GATATCGGCAGTGCGCCTTCCAGGAGAATGATGTCTACGCACAAATACCACTGCGTATGC 335
Qy 271 CCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGGTATGCAGAAAGACT------T 324
Db 336 CCTATGCGCAGCGCAGAAGACCCCGAAAGGCTCGATAGCTACGCAAAGAAACTGGCAGCG 395
Qy 325 AAGAAAAGTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACCTGCTGGAAGTC 384
Db 396 GCCTCCGGGAAGGTGCTGGATAGAGAGATCGCAGGAAAAATCACCGACCTGCAGACCGTC 455
Qy 385 ATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACGCCACGTGTAGG 444
Db 456 ATGGCTACGCCAGACGCTGAATCTCCTACCTTTTGCCTGCATACAGACGTCACGTGTCGT 515
Qy 445 TACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCACCGACATCAATC 504
Db 516 ACGGCAGCCGAAGTGGCCGTATACCAGGACGTGTATGCTGTACATGCACCAACATCGCTG 575
Qy 505 TACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTGACACGACCCCT 564
Db 576 TACCATCAGGCGATGAAAGGTGTCAGAACGGCGTATTGGATTGGGTTTGACACCACCCCG 635
Qy 565 TTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACTGGGCTGACGAG 624
Db 636 TTTATGTTTGACGCGCTAGCAGGCGCGTATCCAACCTACGCCACAAACTGGGCCGACGAG 695
Qy 625 AGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGGAGAGCAGGCTT 684
Db 696 CAGGTGTTACAGGCCAGGAACATAGGACTGTGTGCAGCATCCTTGACTGAGGGAAGACTC 755
Qy 685 GGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGATCATATTCTCG 744
Db 756 GGCAAACTGTCCATTCTCCGCAAGAAGCAATTGAAACCTTGCGACACAGTCATGTTCTCG 815
Qy 745 GTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGCATCTTCCAAAC 804
Db 816 GTAGGATCTACATTGTACACTGAGAGCAGAAAGCTACTGAGGAGCTGGCACTTACCCTCC 875
Qy 805 GTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTGTCAGCTGT 864
Db 876 GTATTCCACCTGAAAGGTAAACAATCCTTTACCTGTAGGTGCGATACCATCGTATCATGT 935
Qy 865 GAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAGTTGAGAAC 924
Db 936 GAAGGGTACGTAGTTAAGAAAATCACTATGTGCCCCGGCCTGTACGGTAAAACGGTAGGG 995
Qy 925 TTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATACGCTGCGC 984
Db 996 TACGCCGTGACGTATCACGCGGAGGGATTCCTAGTGTGCAAGACCACAGACACTGTCAAA 1055
Qy 985 GGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCGATCAGATG 1044
Db 1056 GGAGAAAGAGTCTCATTCCCTGTATGCACCTACGTCCCCTCAACCATCTGTGATCAAATG 1115
Qy 1045 ACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTATTGGTTGGGCTC 1104
Db 1116 ACTGGCATACTAGCGACCGACGTCACACCGGAGGACGCACAGAAGTTGTTAGTGGGATTG 1175
Qy 1105 AACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAATGCAGAACTAT 1164
Db 1176 AATCAGAGGATAGTTGTGAACGGAAGAACACAGCGAAACACTAACACGATGAAGAACTAT 1235
Qy 1165 CTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATCGTGCCGACTTG 1224
Db 1236 CTGCTTCCGATTGTGGCCGTCGCATTTAGCAAGTGGGCGAGGGAATACAAGGCAGACCTT 1295
Qy 1225 GACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCTGCTGCTGGGCT 1284
Db 1296 GATGATGAAAAACCTCTGGGTGTCCGAGAGAGGTCACTTACTTGCTGCTGCTTGTGGGCA 1355
Qy 1285 TTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCTGGTACGCAAACAATTAAGAAA 1344
Db 1356 TTTAAAACGAGGAAGATGCACACCATGTACAAGAAACCAGACACCCAGACAATAGTGAAG 1415
Qy 1345 GTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCACGGGCTCGATATG 1404
Db 1416 GTGCCTTCAGAGTTTAACTCGTTCGTCATCCCGAGCCTATGGTCTACAGGCCTCGCAATC 1475
Qy 1405 GGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACTGTCAAACCCGCACCGGCTATT 1464
Db 1476 CCAGTCAGATCACGCATTAAGATGCTTTTGGCCAAGAAGACCAAGC---GAGAGTTAATA 1532
Qy 1465 ACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAGAAGTGGCTGCA 1524
Db 1533 CCTGTTCTCGACGCGTCGTCAGCCAGGGATGCTGAACAAGAGGAGAAGGAGAGGTTGGAG 1592
Qy 1525 GCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAAAAGAGACCGTA 1584
Db 1593 GCCGAGCTGACTAGAGAAGCCTTACCACCCCTCGTCCCCATCGCGCCGGCGGAGACGGGA 1652
Qy 1585 GAGGCAGAAGTAGACCTCATTATGCAAGAG------GCAGGAGCAGGTAGCGTGGAGACA 1638
Db 1653 GTCGTCGACGTCGACGTTGAAGAACTAGAGTATCACGCAGGTGCAGGGGTCGTGGAAACA 1712
Qy 1639 CCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAGATTGGGTCTTACGCT 1698
Db 1713 CCTCGCAGCGCGTTGAAAGTCACCGCACAGCCGAACGACGTACTACTAGGAAATTACGTA 1772
Qy 1699 ATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGTATCCACCCATTGGCG 1758
Db 1773 GTTCTGTCCCCGCAGACCGTGCTCAAGAGCTCCAAGTTGGCCCCCGTGCACCCTCTAGCA 1832
Qy 1759 GAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATACAAAGTCGAGCCATAC 1818
Db 1833 GAGCAGGTGAAAATAATAACACATAACGGGAGGGCCGGCGGTTACCAGGTCGACGGATAT 1892
Qy 1819 CACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAAGACTTCCAGGCATTG 1878
Db 1893 GACGGCAGGGTCCTACTACCATGTGGATCGGCCATTCCGGTCCCTGAGTTTCAAGCTTTG 1952
Qy 1879 AGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAACAGATACCTGCACCAC 1938
Db 1953 AGCGAGAGCGCCACTATGGTGTACAACGAAAGGGAGTTCGTCAACAGGAAACTATACCAT 2012
Qy 1939 ATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTATAAGACTGTAAAGACT 1998
Db 2013 ATTGCCGTTCACGGACCGTCGCTGAACACCGACGAGGAGAACTACGAGAAAGTCAGAGCT 2072
Qy 1999 CAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAGTGTGTTAAGCGAGAA 2058
Db 2073 GAAAGAACTGACGCCGAGTACGTGTTCGACGTAGATAAAAAATGCTGCGTCAAGAGAGAG 2132
Qy 2059 GACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCATTTCACGAGTTTGCG 2118
Db 2133 GAAGCGTCGGGTTTGGTGTTGGTGGGAGAGCTAACCAACCCCCCGTTCCATGAATTCGCC 2192
Qy 2119 TACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCAACCATCGGAGTCTAT 2178
Db 2193 TACGAAGGGCTGAAGATCAGGCCGTCGGCACCATATAAGACTACAGTAGTAGGAGTCTTT 2252
Qy 2179 GGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTGACTAAGAAAGATCTG 2238
Db 2253 GGGGTTCCGGGATCAGGCAAGTCTGCTATTATTAAGAGCCTCGTGACCAAACACGATCTG 2312
Qy 2239 GTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGATGTAAGGAGGATGAGA 2298
Db 2313 GTCACCAGCGGCAAGAAGGAGAACTGCCAGGAAATAGTTAACGACGTGAAGAAGCACCGC 2372


Qy 2299 CGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAATGGGGTTAAGCACCCC 2358
Db 2373 GGGAAGGGGACAAGTAGGGAAAACAGTGACTCCATCCTGCTAAACGGGTGTCGTCGTGCC 2432
Qy 2359 GTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGGACGCTGCTGGCACTG 2418
Db 2433 GTGGACATCCTATATGTGGACGAGGCTTTCGCTTGCCATTCCGGTACTCTGCTGGCCCTA 2492
Qy 2419 ATTGCCATCGTCAAACCT---AAGAAAGTGGTATTGTGCGGGGACCCAAAACAATGCGGC 2475
Db 2493 ATTGCTCTTGTTAAACCTCGGAGCAAAGTGGTGTTATGCGGAGACCCCAAGCAATGCGGA 2552
Qy 2476 TTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGCACTGAAGTGTAC 2535
Db 2553 TTCTTCAATATGATGCAGCTTAAGGTGAACTTCAACCACAACATCTGCACTGAAGTATGT 2612
Qy 2536 CATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATCGTCTCCACGCTCTTC 2595
Db 2613 CATAAAAGTATATCCAGACGTTGCACGCGTCCAGTCACGGCCATCGTGTCTACGTTGCAC 2672
Qy 2596 TACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATCATCATAGATACCACA 2655
Db 2673 TACGGAGGCAAGATGCGCACGACCAACCCGTGCAACAAACCCATAATCATAGACACCACA 2732
Qy 2656 GGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGAGGATGGGTGAAA 2715
Db 2733 GGACAGACCAAGCCCAAGCCAGGAGACATCGTGTTAACATGCTTCCGAGGCTGGGCAAAG 2792
Qy 2716 CAGCTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCTGCATCGCAAGGACTT 2775
Db 2793 CAGCTGCAGTTGGACTACCGTGGACACGAAGTCATGACAGCAGCAGCATCTCAGGGCCTC 2852
Qy 2776 ACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCACTCTACTCGCAG 2835
Db 2853 ACCCGCAAAGGGGTATACGCCGTAAGGCAGAAGGTGAATGAAAATCCCTTGTATGCCCCT 2912
Qy 2836 ACTTCTGAGCACGTGAACGTGTTACTTACACGCACAGAAAAACGCATTGTCTGGAAGACG 2895
Db 2913 GCGTCGGAGCACGTGAATGTACTGCTGACGCGCACTGAGGATAGGCTGGTGTGGAAAACG 2972
Qy 2896 CTAGCTGGTGATCCCTGGATAAAGACACTTACAGCTAAATATCCCGGGGATTTCACGGCT 2955
Db 2973 CTGGCCGGCGATCCCTGGATTAAGGTCCTATCAAACATTCCACAGGGTAACTTTACGGCC 3032
Qy 2956 TCATTGGACGACTGGCAGCGCGAACACGACGCCATTATGGCACGCGTTCTTGATAAGCCG 3015
Db 3033 ACATTGGAAGAATGGCAAGAAGAACACGACAAAATAATGAAGGTGATTGAAGGACCGGCT 3092
Qy 3016 CAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAGGCTTTAGAGCCA 3075
Db 3093 GCGCCTGTGGACGCGTTCCAGAACAAAGCGAACGTGTGTTGGGCGAAAAGCCTGGTGCCT 3152
Qy 3076 GTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAGAC---GTTGCACCCA 3132
Db 3153 GTCCTGGACACTGCCGGAATCAGATTGACAGCAGAGGAGTGGAGCACCATAATTACAGCA 3212
Qy 3133 TTCAAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTTTGCACCAGGTTC 3192
Db 3213 TTTAAGGAGGACAGAGCTTACTCTCCAGTGGTGGCCTTGAATGAAATTTGCACCAAGTAC 3272
Qy 3193 TTTGGAGTAGACCTGGACAGTGGGTTATTTTCCGCTCCTACCGTCGCACTTACTTACAGG 3252
Db 3273 TATGGAGTTGACCTGGACAGTGGCCTGTTTTCTGCCCCGAAGGTGTCCCTGTATTACGAG 3332
Qy 3253 GATCAGCACTGGGATAACTCGCCAGGGAAGAACATGTATGGGCTTAATAGAGAGGTAGCA 3312
Db 3333 AACAACCACTGGGATAACAGACCTGGTGGAAGGATGTATGGATTCAATGCCGCAACAGCT 3392
Qy 3313 AAGGAGTTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACAGGCAGGGTAGCT 3372
Db 3393 GCCAGGCTGGAAGCTAGACATACCTTCCTGAAGGGGCAGTGGCATACGGGCAAGCAGGCA 3452
Qy 3373 GATATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTGGTTCCATTAAAT 3432
Db 3453 GTTATCGCAGAAAGAAAAATCCAACCGCTTTCTGTGCTGGACAATGTAATTCCTATCAAC 3512
Qy 3433 CGCCGGTTGCCCCACTCGTTGATCGTTGACCACAAAGGACAGGGTACAACTGATCACAGC 3492
Db 3513 CGCAGGCTGCCGCACGCCCTGGTGGCTGAGTACAAGACGGTTAAAGGCAGTAGGGTTGAG 3572
Qy 3493 GGATTCCTATCTAAGATGAAGGGCAAATCTGTGTTGGTGATCGGCGATCCTATCAGCATT 3552
Db 3573 TGGCTGGTCAATAAAGTAAGAGGGTACCACGTCCTGCTGGTGAGTGAGTACAACCTGGCT 3632
Qy 3553 CCAGGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATAC---------CATCAGG 3603
Db 3633 TTGCCTCGACGCAGGGTCACTTGGTTGTCACCGCTGAATGTCACAGGCGCCGATAGGTGC 3692
Qy 3604 TGTGATCTCGATTTGGGAATACCTAGCCATGTCGGTAAATATGACATTATCTTTGTCAAT 3663
Db 3693 TACGACCTAAGTTTAGGACTGCCGGCTGACGCCGGCAGGTTCGACTTGGTCTTTGTGAAC 3752
Qy 3664 GTTAGGACCCCGTACAGGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCACCAC 3723
Db 3753 ATTCACACGGAATTCAGAATCCACCACTACCAGCAGTGTGTCGACCACGCCATGAAGCTG 3812
Qy 3724 AGCATGCTAACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCTATA 3783
Db 3813 CAGATGCTTGGGGGAGATGCGCTACGACTGCTAAAACCCGGCGGCATCT---TGATGAGA 3869
Qy 3784 GGGTATGGGCTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCACGCTCATTT 3843
Db 3870 GCTTACGGATACGCCGATAAAATCAGCGAAGCCGTTGTTTCCTCCTTAAGCAGAAAGTTC 3929
Qy 3844 AGGTTTACCCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTCGTG 3903
Db 3930 TCGTCTGCAAGAGTGTTGCGCCCGGATTGTGTCACCAGCAATACAGAAGTGTTCTTGCTG 3989
Qy 3904 TTCTTCGGCAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTGCTT 3963
Db 3990 TTCTCCAACTTTGACAACGGAAAGAGACCCTCTACGCTACACCAGATGAATACCAAGCTG 4049
Qy 3964 GACAACATCTATCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGTACAGAGTG 4023
Db 4050 AGTGCCGTGTATGCCGGAGAAGCCATGCACACGGCCGGGTGTGCACCATCCTACAGAGTT 4109
Qy 4024 ATCAGAGGTGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGCAAA 4083
Db 4110 AAGAGAGCAGACATAGCCACGTGCACAGAAGCGGCTGTGGTTAACGCAGCTAACGCCCGT 4169
Qy 4084 GGTCAACCAGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTTGAT 4143
Db 4170 GGAACTGTAGGGGATGGCGTATGCAGGGCCGTGGCGAAGAAATGGCCGTCAGCCTTTAAG 4229
Qy 4144 AGACAGCCAATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACCG---CTCATCATA 4200
Db 4230 GGAGCAGCAACACCAGTGGGCACAATTAAAACAGTCATGTGCGGCTCGTACCCCGTCATC 4289
Qy 4201 CATGCTGTAGGACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTAAGCTCGCA 4260
Db 4290 CACGCTGTAGCGCCTAATTTCTCTGCCACGACTGAAGCGGAAGGGGACCGCGAATTGGCC 4349
Qy 4261 GCTGCCTACATGAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTA 4320
Db 4350 GCTGTCTACCGGGCAGTGGCCGCCGAAGTAAACAGACTGTCACTGAGCAGCGTAGCCATC 4409
Qy 4321 CCGCTACTGTCAACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCAT 4380
Db 4410 CCGCTGCTGTCCACAGGAGTGTTCAGCGGCGGAAGAGATAGGCTGCAGCAATCCCTCAAC 4469
Qy 4381 CACCTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAA 4440
Db 4470 CATCTATTCACAGCAATGGACGCCACGGACGCTGACGTGACCATCTACTGCAGAGACAAA 4529
Qy 4441 CAATGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATTCTGGAT 4500

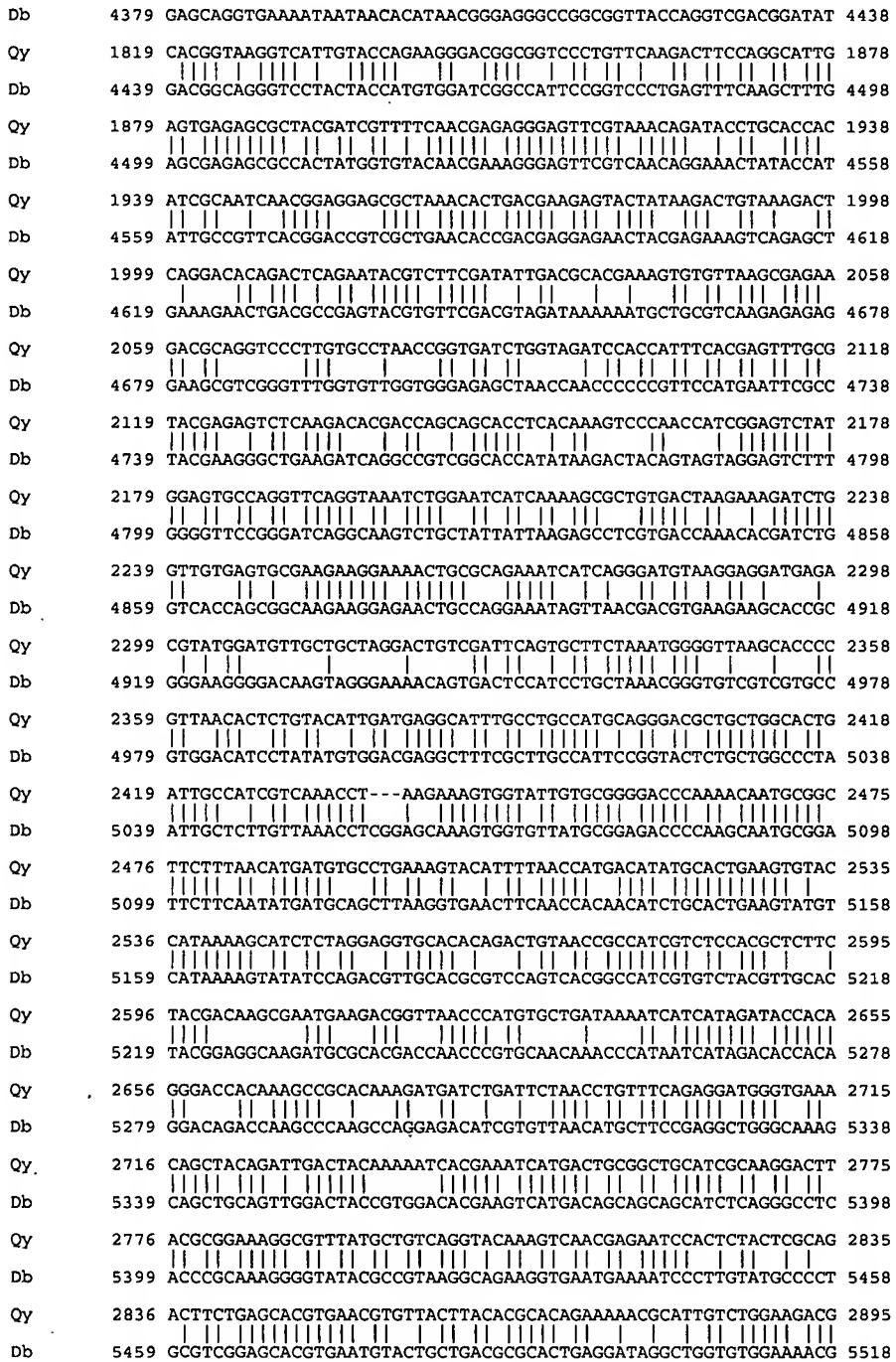
```
Db       4530 AGTTGGGAGAAGAAAATCCAGGAAGCCATTGACATGAGGACGGCTGTGGAGTTGCTCAAT 4589

Qy       4501 GATGACAAGCCAGTAGACATTGACTTGGTCAGGGTCCACCCAAACAGCTCTTTGGCAGGC 4560

Db       4590 GATGACGTGGAGCTGACCACAGACTTGGTGAGAGTGCACCCGGACAGCAGCCTGGTGGGT 4649

Qy       4561 AGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTGGAAGGTACACGATTC 4620

Db       4650 CGTAAGGGCTACAGTACCACTGACGGGTCGCTGTACTCGTACTTTGAAGGTACGAAATTC 4709

Qy       4621 CATCAGACCGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCCAACAAATCTGAGGCT 4680

Db       4710 AACCAGGCTGCTATTGATATGGCAGAGATACTGACGTTGTGGCCCAGACTGCAAGAGGCA 4769

Qy       4681 AATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGC 4740

Db       4770 AACGAACAGATATGCCTATACGCGCTGGGCGAAACAATGGACAACATCAGATCCAAATGT 4829

Qy       4741 CCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTAC 4800

Db       4830 CCGGTGAACGATTCCGATTCATCAACACCTCCCAGGACAGTGCCCTGCCTGTGCCGCTAC 4889

Qy       4801 GCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTA 4860

Db       4890 GCAATGACAGCAGAACGGATCGCCCGCCTTAGGTCACACCAAGTTAAAAGCATGGTGGTT 4949

Qy       4861 TGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGC 4920

Db       4950 TGCTCATCTTTTCCCCTCCCGAAATACCATGTAGATGGGGTGCAGAAGGTAAAGTGCGAG 5009

Qy       4921 AAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGCGGAA 4980

Db       5010 AAGGTTCTCCTGTTCGACCCGACGGTACCTTCAGTGGTTAGTCCGCGGAAGTATGCCGCA 5069


RESULT 5
US-08-466-277-1
; Sequence 1, Application US/08466277
; Patent No. 6190666
;   GENERAL INFORMATION:
;        APPLICANT: Garoff, Henrik
;                   Liljestrom, Peter
;        TITLE OF INVENTION: DNA Expression Systems Based on
;                            Alphaviruses
;        NUMBER OF SEQUENCES: 27
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Birch, Stewart, Kolasch & Birch
;             STREET: P.O. Box 747
;             CITY: Falls Church
;             STATE: Virginia
;             COUNTRY: USA
;             ZIP: 22040-0747
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/466,277
;             FILING DATE: 06-Jun-1995
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 07/920,281
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Murphy Jr., Gerald M.
;             REGISTRATION NUMBER: 28,977
;             REFERENCE/DOCKET NUMBER: 828-103P
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 703-241-1300
;             TELEFAX: 703-241-2848
;             TELEX: 248345
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 11517 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: RNA (genomic)
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        ORIGINAL SOURCE:
;             ORGANISM: Semliki Forest Virus
;        FEATURE:
;             NAME/KEY:  -
;             LOCATION:  1..11517
;             OTHER INFORMATION:  /label= genome
;             /note= "Semliki Forest Virus complete nucleotide
;             sequence, presented as a cloned DNA sequence; see
;             Figure 5."
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  87..7379
;             OTHER INFORMATION:  /product= "SFV polyprotein"
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  7421..11179
;             OTHER INFORMATION:  /product= "SFV polyprotein"
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-08-466-277-1

  Query Match             31.3%;  Score 1564;  DB 3;  Length 11517;
  Best Local Similarity   58.3%;  Pred. No. 0;
  Matches 2904;  Conservative    0;  Mismatches 2040;  Indels   36;  Gaps    8;

Qy         31 AGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGTCGTTACAGCGGACGTTT 90

Db         96 AAAGTGCATGTTGATATTGAGGCTGACAGCCCATTCATCAAGTCTTTGCAGAAGGCATTT 155

Qy         91 CCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACCATGCCAATGCCAGAGCG 150

Db        156 CCGTCGTTCGAGGTGGAGTCATTGCAGGTCACACCAAATGACCATGCAAATGCCAGAGCA 215

Qy        151 TTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACCGGGACCAAGTTATCTTG 210

Db        216 TTTTCGCACCTGGCTACCAAATTGATCGAGCAGGAGACTGACAAAGACACACTCATCTTG 275

Qy        211 GACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACCGCTATCATTGTATCTGC 270

Db        276 GATATCGGCAGTGCGCCTTCCAGGAGAATGATGTCTACGCACAAATACCACTGCGTATGC 335

Qy        271 CCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGGTATGCAGAAAGACT------T 324

Db        336 CCTATGCGCAGCGCAGAAGACCCCGAAAGGCTCGATAGCTACGCAAAGAAACTGGCAGCG 395

Qy        325 AAGAAAAGTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACCTGCTGGAAGTC 384

Db        396 GCCTCCGGGAAGGTGCTGGATAGAGAGATCGCAGGAAAAATCACCGACCTGCAGACCGTC 455

Qy        385 ATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACGCCACGTGTAGG 444

Db        456 ATGGCTACGCCAGACGCTGAATCTCCTACCTTTGCCTGCATACAGACGTCACGTGTCGT 515

Qy        445 TACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCACCGACATCAATC 504

Db        516 ACGGCAGCCGAAGTGGCCGTATACCAGGACGTGTATGCTGTACATGCACCAACATCGCTG 575

Qy        505 TACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTGACACGACCCCT 564

Db        576 TACCATCAGGCGATGAAAGGTGTCAGAACGGCGTATTGGATTGGGTTTGACACCACCCCG 635

Qy        565 TTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACTGGGCTGACGAG 624

Db        636 TTTATGTTTGACGCGCTAGCAGGCGCGTATCCAACCTACGCCACAAACTGGGCCGACGAG 695

Qy        625 AGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGGAGAGCAGGCTT 684
```

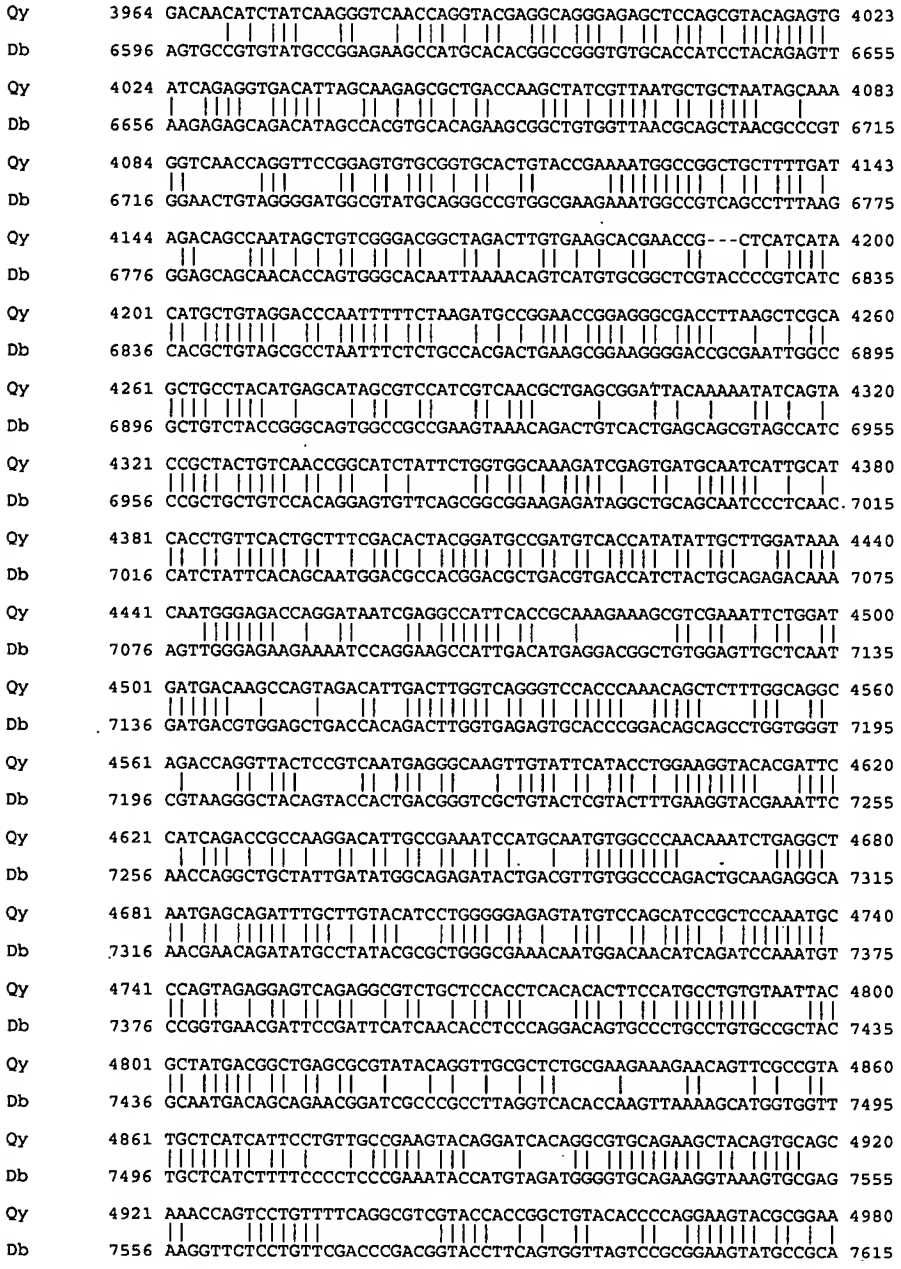

Db 696 CAGGTGTTACAGGCCAGGAACATAGGACTGTGTGCAGCATCCTTGACTGAGGGAAGACTC 755
Qy 685 GGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGATCATATTCTCG 744
Db 756 GGCAAACTGTCCATTCTCCGCAAGAAGCAATTGAAACCTTGCGACACAGTCATGTTCTCG 815
Qy 745 GTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGCATCTTCCAAAC 804
Db 816 GTAGGATCTACATTGTACACTGAGAGCAGAAAGCTACTGAGGAGCTGGCACTTACCCTCC 875
Qy 805 GTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTGTCAGCTGT 864
Db 876 GTATTCCACCTGAAAGGTAAACAATCCTTTACCTGTAGGTGCGATACCATCGTATCATGT 935
Qy 865 GAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAGTTGAGAAC 924
Db 936 GAAGGGTACGTAGTTAAGAAAATCACTATGTGCCCCGGCCTGTACGGTAAAACGGTAGGG 995
Qy 925 TTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATACGCTGCGC 984
Db 996 TACGCCGTGACGTATCACGCGGAGGGATTCCTAGTGTGCAAGACCACAGACACTGTCAAA 1055
Qy 985 GGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCGATCAGATG 1044
Db 1056 GGAGAAAGAGTCTCATTCCCTGTATGCACCTACGTCCCCTCAACCATCTGTGATCAAATG 1115
Qy 1045 ACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTATTGGTTGGGCTC 1104
Db 1116 ACTGGCATACTAGCGACCGACGTCACACCGGAGGACGCACAGAAGTTGTTAGTGGGATTG 1175
Qy 1105 AACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAATGCAGAACTAT 1164
Db 1176 AATCAGAGGATAGTTGTGAACGGAAGAACACAGCGAAACACTAACACGATGAAGAACTAT 1235
Qy 1165 CTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATCGTGCCGACTTG 1224
Db 1236 CTGCTTCCGATTGTGGCCGTCGCATTTAGCAAGTGGGCGAGGGAATACAAGGCAGACCTT 1295
Qy 1225 GACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCTGCTGCTGGGCT 1284
Db 1296 GATGATGAAAAACCTCTGGGTGTCCGAGAGAGGTCACTTACTTGCTGCTGCTTGTGGGCA 1355
Qy 1285 TTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCTGGTACGCAAACAATTAAGAAA 1344
Db 1356 TTTAAAACGAGGAAGATGCACACCATGTACAAGAAACCAGACACCCAGACAATAGTGAAG 1415
Qy 1345 GTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCACGGGCTCGATATG 1404
Db 1416 GTGCCTTCAGAGTTTAACTCGTTCGTCATCCCGAGCCTATGGTCTACAGGCCTCGCAATC 1475
Qy 1405 GGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACTGTCAAACCCGCACCGGCTATT 1464
Db 1476 CCAGTCAGATCACGCATTAAGATGCTTTTGGCCAAGAAGACCAAGC---GAGAGTTAATA 1532
Qy 1465 ACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAGAAGTGGCTGCA 1524
Db 1533 CCTGTTCTCGACGCGTCGTCAGCCAGGGATGCTGAACAAGAGGAGAAGGAGAGGTTGGAG 1592
Qy 1525 GCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAAAAGAGACCGTA 1584
Db 1593 GCCGAGCTGACTAGAGAAGCCTTACCACCCCTCGTCCCCATCGCGCCGGCGGAGACGGGA 1652
Qy 1585 GAGGCAGAAGTAGACCTCATTATGCAAGAG------GCAGGAGCAGGTAGCGTGGAGACA 1638
Db 1653 GTCGTCGACGTCGACGTTGAAGAACTAGAGTATCACGCAGGTGCAGGGGTCGTGGAAACA 1712
Qy 1639 CCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAGATTGGGTCTTACGCT 1698
Db 1713 CCTCGCAGCGCGTTGAAAGTCACCGCACAGCCGAACGACGTACTACTAGGAAATTACGTA 1772
Qy 1699 ATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGTATCCACCCATTGGCG 1758
Db 1773 GTTCTGTCCCCGCAGACCGTGCTCAAGAGCTCCAAGTTGGCCCCCGTGCACCCTCTAGCA 1832
Qy 1759 GAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATACAAAGTCGAGCCATAC 1818
Db 1833 GAGCAGGTGAAAATAATAACACATAACGGGAGGGCCGGCGGTTACCAGGTCGACGGATAT 1892
Qy 1819 CACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAAGACTTCCAGGCATTG 1878
Db 1893 GACGGCAGGGTCCTACTACCATGTGGATCGGCCATTCCGGTCCCTGAGTTTCAAGCTTTG 1952
Qy 1879 AGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAACAGATACCTGCACCAC 1938
Db 1953 AGCGAGAGCGCCACTATGGTGTACAACGAAAGGGAGTTCGTCAACAGGAAACTATACCAT 2012
Qy 1939 ATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTATAAGACTGTAAAGACT 1998
Db 2013 ATTGCCGTTCACGGACCGTCGCTGAACACCGACGAGGAGAACTACGAGAAAGTCAGAGCT 2072
Qy 1999 CAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAGTGTGTTAAGCGAGAA 2058
Db 2073 GAAAGAACTGACGCCGAGTACGTGTTCGACGTAGATAAAAAATGCTGCGTCAAGAGAGAG 2132
Qy 2059 GACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCATTTCACGAGTTTGCG 2118
Db 2133 GAAGCGTCGGGTTTGGTGTTGGTGGGAGAGCTAACCAACCCCCCGTTCCATGAATTCGCC 2192
Qy 2119 TACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCAACCATCGGAGTCTAT 2178
Db 2193 TACGAAGGGCTGAAGATCAGGCCGTCGGCACCATATAAGACTACAGTAGTAGGAGTCTTT 2252
Qy 2179 GGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTGACTAAGAAAGATCTG 2238
Db 2253 GGGGTTCCGGGATCAGGCAAGTCTGCTATTATTAAGAGCCTCGTGACCAAACACGATCTG 2312
Qy 2239 GTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGATGTAAGGAGGATGAGA 2298
Db 2313 GTCACCAGCGGCAAGAAGGAGAACTGCCAGGAAATAGTTAACGACGTGAAGAAGCACCGC 2372
Qy 2299 CGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAATGGGGTTAAGCACCCC 2358
Db 2373 GGGAAGGGGACAAGTAGGGAAAACAGTGACTCCATCCTGCTAAACGGGTGTCGTCGTGCC 2432
Qy 2359 GTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGGACGCTGCTGGCACTG 2418
Db 2433 GTGGACATCCTATATGTGGACGAGGCTTTCGCTTGCCATTCCGGTACTCTGCTGGCCCTA 2492
Qy 2419 ATTGCCATCGTCAAACCT---AAGAAAGTGGTATTGTGCGGGGACCCAAAACAATGCGGC 2475
Db 2493 ATTGCTCTTGTTAAACCTCGGAGCAAAGTGGTGTTATGCGGAGACCCCAAGCAATGCGGA 2552
Qy 2476 TTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGCACTGAAGTGTAC 2535
Db 2553 TTCTTCAATATGATGCAGCTTAAGGTGAACTTCAACCACAACATCTGCACTGAAGTATGT 2612
Qy 2536 CATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATCGTCTCCACGCTCTTC 2595
Db 2613 CATAAAAGTATATCCAGACGTTGCACGCGTCCAGTCACGGCCATCGTGTCTACGTTGCAC 2672
Qy 2596 TACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATCATCATAGATACCACA 2655
Db 2673 TACGGAGGCAAGATGCGCACGACCAACCCGTGCAACAAACCCATAATCATAGACACCACA 2732
Qy 2656 GGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGAGGATGGGTGAAA 2715
Db 2733 GGACAGACCAAGCCCAAGCCAGGAGACATCGTGTTAACATGCTTCCGAGGCTGGGCAAAG 2792
Qy 2716 CAGCTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCTGCATCGCAAGGACTT 2775
Db 2793 CAGCTGCAGTTGGACTACCGTGGACACGAAGTCATGACAGCAGCAGCATCTCAGGGCCTC 2852
Qy 2776 ACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCACTCTACTCGCAG 2835
Db 2853 ACCCGCAAAGGGGTATACGCCGTAAGGCAGAAGGTGAATGAAAATCCCTTGTATGCCCCT 2912





```
RESULT 6
US-09-554-337-1
; Sequence 1, Application US/09554337
; Patent No. 6475780
; GENERAL INFORMATION:
;  APPLICANT: Parrington, Mark
;  APPLICANT:  Li, Xiaomao
;  APPLICANT:  Klein, Michel H.
;  TITLE OF INVENTION: ALPHAVIRUS VECTORS FOR PARAMYXOVIRUS VACCINES
;  FILE REFERENCE: 1038-1042 MIS
;  CURRENT APPLICATION NUMBER: US/09/554,337
;  CURRENT FILING DATE:  2000-10-20
;  PRIOR APPLICATION NUMBER: 60/065,791
;  PRIOR FILING DATE: 1997-11-14
;  PRIOR APPLICATION NUMBER: PCT/CA98/01064
;  PRIOR FILING DATE: 1998-11-13
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 15538
;   TYPE: DNA
;   ORGANISM: respiratory syncytial virus
US-09-554-337-1

  Query Match             31.3%;  Score 1564;  DB 4;  Length 15538;
  Best Local Similarity   58.3%;  Pred. No. 0;
  Matches 2904;  Conservative    0;  Mismatches 2040;  Indels   36;  Gaps    8;

Qy         31 AGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGTCGTTACAGCGGACGTTT 90
Db       2642 AAAGTGCATGTTGATATTGAGGCTGACAGCCCATTCATCAAGTCTTTGCAGAAGGCATTT 2701

Qy         91 CCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACCATGCCAATGCCAGAGCG 150
Db       2702 CCGTCGTTCGAGGTGGAGTCATTGCAGGTCACACCAAATGACCATGCAAATGCCAGAGCA 2761

Qy        151 TTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACCGGGACCAAGTTATCTTG 210
Db       2762 TTTTCGCACCTGGCTACCAAATTGATCGAGCAGGAGACTGACAAAGACACACTCATCTTG 2821

Qy        211 GACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACCGCTATCATTGTATCTGC 270
Db       2822 GATATCGGCAGTGCGCCTTCCAGGAGAATGATGTCTACGCACAAATACCACTGCGTATGC 2881

Qy        271 CCTATGATAAGCGCTGAAGACCCCGGACAGACTACAACGGTATGCAGAAAGACT------T 324
Db       2882 CCTATGCGCAGCGCAGAAGACCCCGAAAGGCTCGATAGCTACGCAAAGAAACTGGCAGCG 2941

Qy        325 AAGAAAAGTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACCTGCTGGAAGTC 384
Db       2942 GCCTCCGGGAAGGTGCTGGATAGAGAGATCGCAGGAAAAATCACCGACCTGCAGACCGTC 3001

Qy        385 ATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACGCCACGTGTAGG 444
Db       3002 ATGGCTACGCCAGACGCTGAATCTCCTACCTTTTGCCTGCATACAGACGTCACGTGTCGT 3061

Qy        445 TACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCACCGACATCAATC 504
Db       3062 ACGGCAGCCGAAGTGGCCGTATACCAGGACGTGTATGCTGTACATGCACCAACATCGCTG 3121

Qy        505 TACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTGACACGACCCCT 564
Db       3122 TACCATCAGGCGATGAAAGGTGTCAGAACGGCGTATTGGATTGGGTTTGACACCACCCCG 3181

Qy        565 TTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACTGGGCTGACGAG 624
Db       3182 TTTATGTTTGACGCGCTAGCAGGCGCGTATCCAACCTACGCCACAAACTGGGCCGACGAG 3241

Qy        625 AGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGGAGAGCAGGCTT 684
Db       3242 CAGGTGTTACAGGCCAGGAACATAGGACTGTGTGCAGCATCCTTGACTGAGGGAAGACTC 3301

Qy        685 GGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGATCATATTCTCG 744

Db       3302 GGCAAACTGTCCATTCTCCGCAAGAAGCAATTGAAACCTTGCGACACAGTCATGTTCTCG 3361

Qy        745 GTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGCATCTTCCAAAC 804
Db       3362 GTAGGATCTACATTGTACACTGAGAGCAGAAAGCTACTGAGGAGCTGGCACTTACCCTCC 3421

Qy        805 GTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTGTCAGCTGT 864
Db       3422 GTATTCCACCTGAAAGGTAAACAATCCTTTACCTGTAGGTGCGATACCATCGTATCATGT 3481

Qy        865 GAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAGTTGAGAAC 924
Db       3482 GAAGGGTACGTAGTTAAGAAAATCACTATGTGCCCCGGCCTGTACGGTAAAACGGTAGGG 3541

Qy        925 TTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATACGCTGCGC 984
Db       3542 TACGCCGTGACGTATCACGCGGAGGGATTCCTAGTGTGCAAGACCACAGACACTGTCAAA 3601

Qy        985 GGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCGATCAGATG 1044
Db       3602 GGAGAAAGAGTCTCATTCCCTGTATGCACCTACGTCCCCTCAACCATCTGTGATCAAATG 3661

Qy       1045 ACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTATTGGTTGGGCTC 1104
Db       3662 ACTGGCATACTAGCGACCGACGTCACACCGGAGGACGCACAGAAGTTGTTAGTGGGATTG 3721

Qy       1105 AACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAATGCAGAACTAT 1164
Db       3722 AATCAGAGGATAGTTGTGAACGGAAGAACACAGCGAAACACTAACACGATGAAGAACTAT 3781

Qy       1165 CTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATCGTGCCGACTTG 1224
Db       3782 CTGCTTCCGATTGTGGCCGTCGCATTTAGCAAGTGGGCGAGGGAATACAAGGCAGACCTT 3841

Qy       1225 GACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCTGCTGCTGGGCT 1284
Db       3842 GATGATGAAAAACCTCTGGGTGTCCGAGAGAGGTCACTTACTTGCTGCTGCTTGTGGGCA 3901

Qy       1285 TTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCTGGTACGCAAACAATTAAGAAA 1344
Db       3902 TTTAAAACGAGGAAGATGCACACCATGTACAAGAAACCAGACACCCAGACAATAGTGAAG 3961

Qy       1345 GTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCACGGGCTCGATATG 1404
Db       3962 GTGCCTTCAGAGTTTAACTCGTTCGTCATCCCGAGCCTATGGTCTACAGGCCTCGCAATC 4021

Qy       1405 GGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACTGTCAAACCCGCACCGGCTATT 1464
Db       4022 CCAGTCAGATCACGCATTAAGATGCTTTTGGCCAAGAAGACCAAGC---GAGAGTTAATA 4078

Qy       1465 ACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAGAAGTGGCTGCA 1524
Db       4079 CCTGTTCTCGACGCGTCGTCAGCCAGGGATGCTGAACAAGAGGAGAAGGAGAGGTTGGAG 4138

Qy       1525 GCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAAAAGAGACCGTA 1584
Db       4139 GCCGAGCTGACTAGAGAAGCCTTACCACCCCTCGTCCCCATCGCGCCGGCGGAGACGGGA 4198

Qy       1585 GAGGCAGAAGTAGACCTCATTATGCAAGAG------GCAGGAGCAGGTAGCGTGGAGACA 1638
Db       4199 GTCGTCGACGTCGACGTTGAAGAACTAGAGTATCACGCAGGTGCAGGGGTCGTGGAAACA 4258

Qy       1639 CCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAGATTGGGTCTTACGCT 1698
Db       4259 CCTCGCAGCGCGTTGAAAGTCACCGCACAGCCGAACGACGTACTACTAGGAAATTACGTA 4318

Qy       1699 ATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGTATCCACCCATTGGCG 1758
Db       4319 GTTCTGTCCCCGCAGACCGTGCTCAAGAGCTCCAAGTTGGCCCCCGTGCACCCTCTAGCA 4378

Qy       1759 GAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATACAAAGTCGAGCCATAC 1818
```

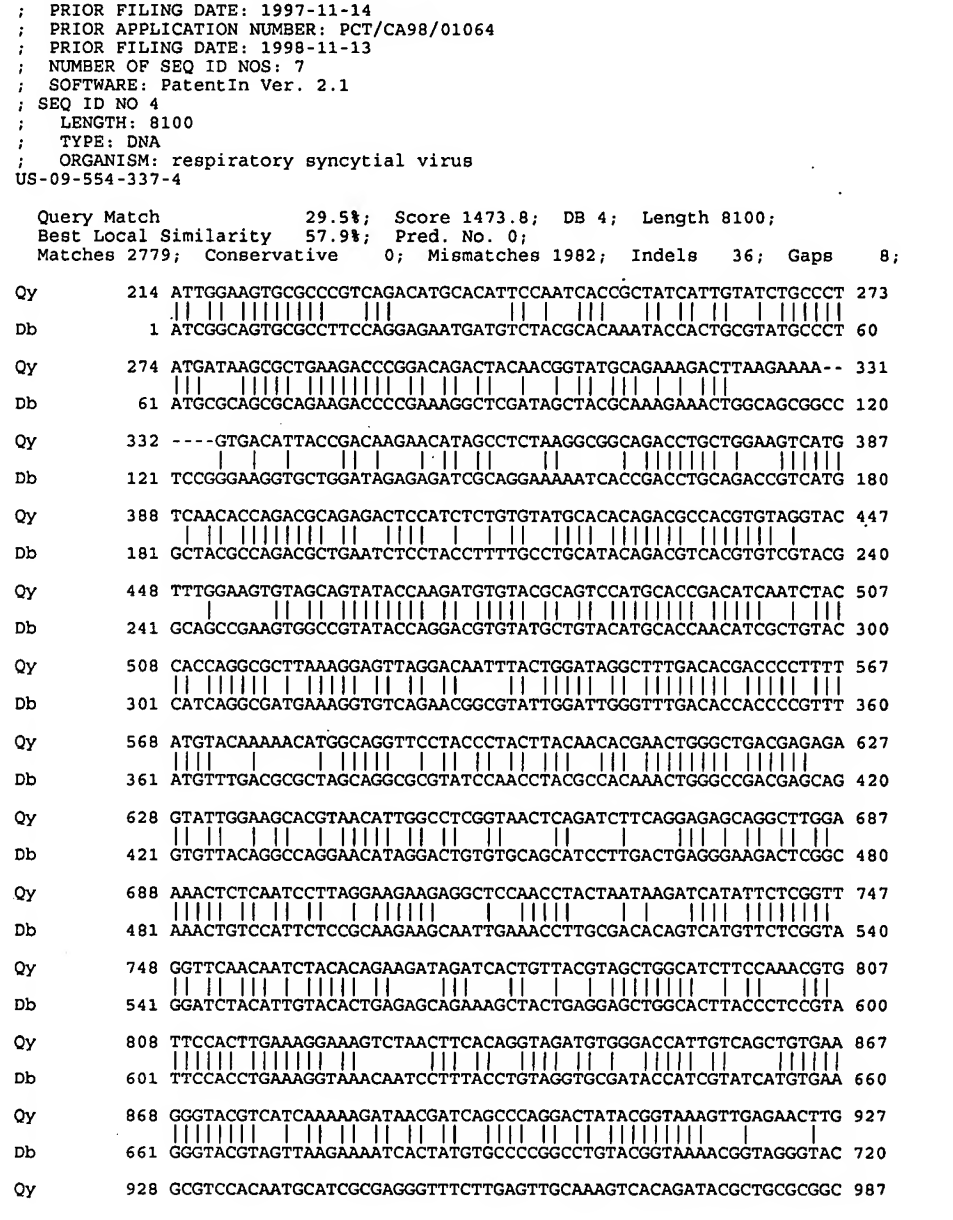

Db 4379 GAGCAGGTGAAAATAATAACACATAACGGGAGGGCCGGCGGTTACCAGGTCGACGGATAT 4438
Qy 1819 CACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAAGACTTCCAGGCATTG 1878
Db 4439 GACGGCAGGGTCCTACTACCATGTGGATCGGCCATTCCGGTCCCTGAGTTTCAAGCTTTG 4498
Qy 1879 AGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAACAGATACCTGCACCAC 1938
Db 4499 AGCGAGAGCGCCACTATGGTGTACAACGAAAGGGAGTTCGTCAACAGGAAACTATACCAT 4558
Qy 1939 ATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTATAAGACTGTAAAGACT 1998
Db 4559 ATTGCCGTTCACGGACCGTCGCTGAACACCGACGAGGAGAACTACGAGAAAGTCAGAGCT 4618
Qy 1999 CAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAGTGTGTTAAGCGAGAA 2058
Db 4619 GAAAGAACTGACGCCGAGTACGTGTTCGACGTAGATAAAAAATGCTGCGTCAAGAGAGAG 4678
Qy 2059 GACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCATTTCACGAGTTTGCG 2118
Db 4679 GAAGCGTCGGGTTTGGTGTTGGTGGGAGAGCTAACCAACCCCCCGTTCCATGAATTCGCC 4738
Qy 2119 TACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCAACCATCGGAGTCTAT 2178
Db 4739 TACGAAGGGCTGAAGATCAGGCCGTCGGCACCATATAAGACTACAGTAGTAGGAGTCTTT 4798
Qy 2179 GGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTGACTAAGAAAGATCTG 2238
Db 4799 GGGGTTCCGGGATCAGGCAAGTCTGCTATTATTAAGAGCCTCGTGACCAAACACGATCTG 4858
Qy 2239 GTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGATGTAAGGAGGATGAGA 2298
Db 4859 GTCACCAGCGGCAAGAAGGAGAACTGCCAGGAAATAGTTAACGACGTGAAGAAGCACCGC 4918
Qy 2299 CGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAATGGGGTTAAGCACCCC 2358
Db 4919 GGGAAGGGGACAAGTAGGGAAAACAGTGACTCCATCCTGCTAAACGGGTGTCGTCGTGCC 4978
Qy 2359 GTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGGACGCTGCTGGCACTG 2418
Db 4979 GTGGACATCCTATATGTGGACGAGGCTTTCGCTTGCCATTCCGGTACTCTGCTGGCCCTA 5038
Qy 2419 ATTGCCATCGTCAAACCT---AAGAAAGTGGTATTGTGCGGGGACCCAAAACAATGCGGC 2475
Db 5039 ATTGCTCTTGTTAAACCTCGGAGCAAAGTGGTGTTATGCGGAGACCCCAAGCAATGCGGA 5098
Qy 2476 TTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGCACTGAAGTGTAC 2535
Db 5099 TTCTTCAATATGATGCAGCTTAAGGTGAACTTCAACCACAACATCTGCACTGAAGTATGT 5158
Qy 2536 CATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATCGTCTCCACGCTCTTC 2595
Db 5159 CATAAAAGTATATCCAGACGTTGCACGCGTCCAGTCACGGCCATCGTGTCTACGTTGCAC 5218
Qy 2596 TACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATCATCATAGATACCACA 2655
Db 5219 TACGGAGGCAAGATGCGCACGACCAACCCGTGCAACAAACCCATAATCATAGACACCACA 5278
Qy 2656 GGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGAGGATGGGTGAAA 2715
Db 5279 GGACAGACCAAGCCCAAGCCAGGAGACATCGTGTTAACATGCTTCCGAGGCTGGGCAAAG 5338
Qy 2716 CAGCTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCTGCATCGCAAGGACTT 2775
Db 5339 CAGCTGCAGTTGGACTACCGTGGACACGAAGTCATGACAGCAGCAGCATCTCAGGGCCTC 5398
Qy 2776 ACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCACTCTACTCGCAG 2835
Db 5399 ACCCGCAAAGGGGTATACGCCGTAAGGCAGAAGGTGAATGAAAATCCCTTGTATGCCCCT 5458
Qy 2836 ACTTCTGAGCACGTGAACGTGTTACTTACACGCACAGAAAAACGCATTGTCTGGAAGACG 2895
Db 5459 GCGTCGGAGCACGTGAATGTACTGCTGACGCGCACTGAGGATAGGCTGGTGTGGAAAACG 5518
Qy 2896 CTAGCTGGTGATCCCTGGATAAAGACACTTACAGCTAAATATCCCGGGGATTTCACGGCT 2955
Db 5519 CTGGCCGGCGATCCCTGGATTAAGGTCCTATCAAACATTCCACAGGGTAACTTTACGGCC 5578
Qy 2956 TCATTGGACGACTGGCAGCGCGAACACGACGCCATTATGGCACGCGTTCTTGATAAGCCG 3015
Db 5579 ACATTGGAAGAATGGCAAGAAGAACACGACAAAATAATGAAGGTGATTGAAGGACCGGCT 5638
Qy 3016 CAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAGGCTTTAGAGCCA 3075
Db 5639 GCGCCTGTGGACGCGTTCCAGAACAAAGCGAACGTGTGTTGGGCGAAAAGCCTGGTGCCT 5698
Qy 3076 GTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAGAC---GTTGCACCCA 3132
Db 5699 GTCCTGGACACTGCCGGAATCAGATTGACAGCAGAGGAGTGGAGCACCATAATTACAGCA 5758
Qy 3133 TTCAAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTTTGCACCAGGTTC 3192
Db 5759 TTTAAGGAGGACAGAGCTTACTCTCCAGTGGTGGCCTTGAATGAAATTTGCACCAAGTAC 5818
Qy 3193 TTTGGAGTAGACCTGGACAGTGGGTTATTTTCCGCTCCTACCGTCGCACTTACTTACAGG 3252
Db 5819 TATGGAGTTGACCTGGACAGTGGCCTGTTTTCTGCCCCGAAGGTGTCCCTGTATTACGAG 5878
Qy 3253 GATCAGCACTGGGATAACTCGCCAGGGAAGAACATGTATGGGCTTAATAGAGAGGTAGCA 3312
Db 5879 AACAACCACTGGGATAACAGACCTGGTGGAAGGATGTATGGATTCAATGCCGCAACAGCT 5938
Qy 3313 AAGGAGTTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACAGGCAGGGTAGCT 3372
Db 5939 GCCAGGCTGGAAGCTAGACATACCTTCCTGAAGGGGCAGTGGCATACGGGCAAGCAGGCA 5998
Qy 3373 GATATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTGGTTCCATTAAAT 3432
Db 5999 GTTATCGCAGAAAGAAAAATCCAACCGCTTTCTGTGCTGGACAATGTAATTCCTATCAAC 6058
Qy 3433 CGCCGGTTGCCCCACTCGTTGATCGTTGACCACAAAGGACAGGGTACAACTGATCACAGC 3492
Db 6059 CGCAGGCTGCCGCACGCCCTGGTGGCTGAGTACAAGACGGTTAAAGGCAGTAGGGTTGAG 6118
Qy 3493 GGATTCCTATCTAAGATGAAGGGCAAATCTGTGTTGGTGATCGGCGATCCTATCAGCATT 3552
Db 6119 TGGCTGGTCAATAAAGTAAGAGGGTACCACGTCCTGCTGGTGAGTGAGTACAACCTGGCT 6178
Qy 3553 CCAGGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATAC---------CATCAGG 3603
Db 6179 TTGCCTCGACGCAGGGTCACTTGGTTGTCACCGCTGAATGTCACAGGCGCCGATAGGTGC 6238
Qy 3604 TGTGATCTCGATTTGGGAATACCTAGCCATGTCGGTAAATATGACATTATCTTTGTCAAT 3663
Db 6239 TACGACCTAAGTTTAGGACTGCCGGCTGACGCCGGCAGGTTCGACTTGGTCTTTGTGAAC 6298
Qy 3664 GTTAGGACCCCGTACAGGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCACCAC 3723
Db 6299 ATTCACACGGAATTCAGAATCCACCACTACCAGCAGTGTGTCGACCACGCCATGAAGCTG 6358
Qy 3724 AGCATGCTAACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCTATA 3783
Db 6359 CAGATGCTTGGGGGAGATGCGCTACGACTGCTAAAACCCGGCGGCATCT---TGATGAGA 6415
Qy 3784 GGGTATGGGCTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCACGCTCATTT 3843
Db 6416 GCTTACGGATACGCCGATAAAATCAGCGAAGCCGTTGTTTCCTCCTTAAGCAGAAAGTTC 6475
Qy 3844 AGGTTTACCCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTCGTG 3903
Db 6476 TCGTCTGCAAGAGTGTTGCGCCCGGATTGTGTCACCAGCAATACAGAAGTGTTCTTGCTG 6535
Qy 3904 TTCTTCGGCAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTGCTT 3963
Db 6536 TTCTCCAACTTTGACAACGGAAAGAGACCCTCTACGCTACACCAGATGAATACCAAGCTG 6595

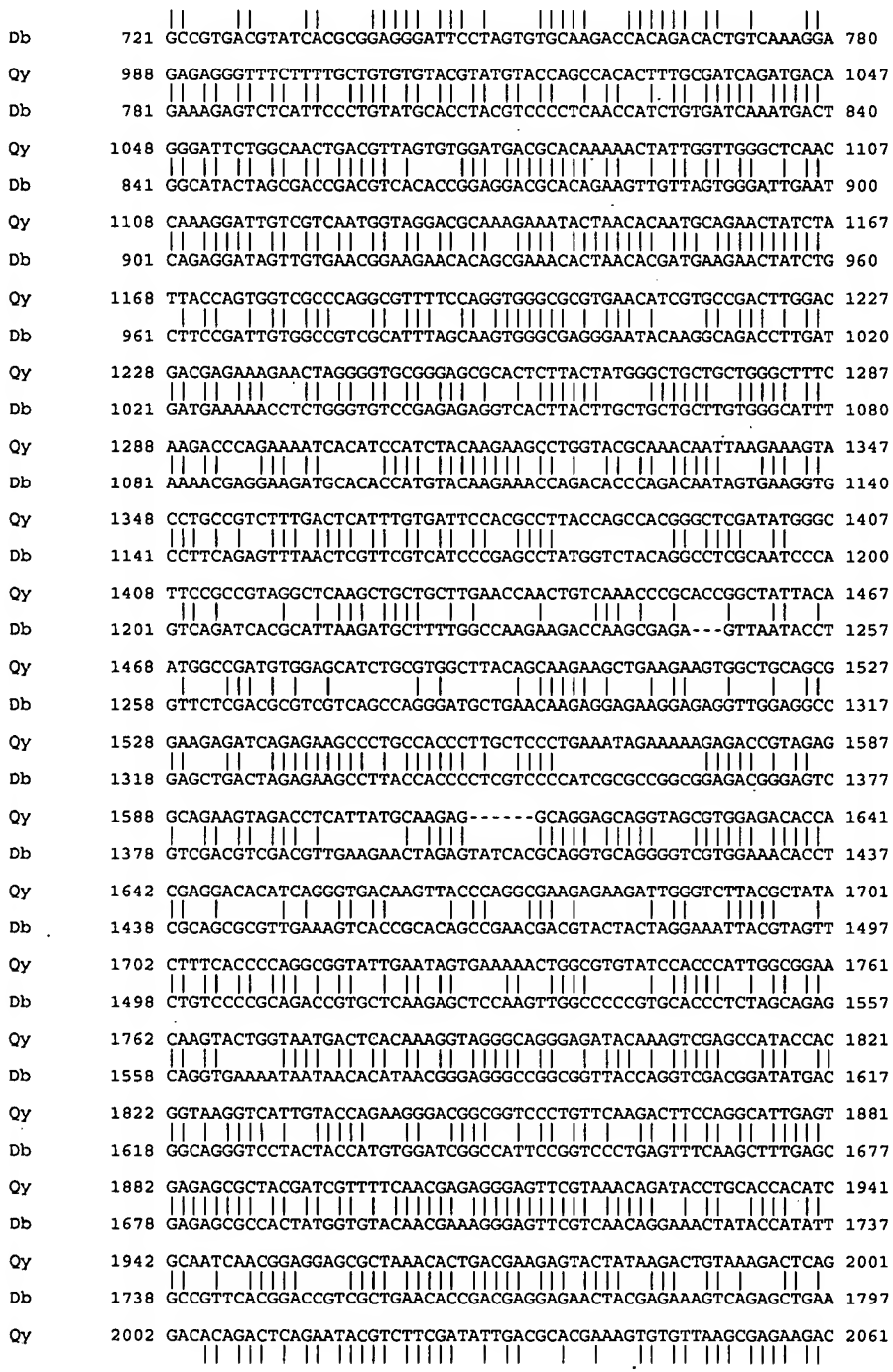


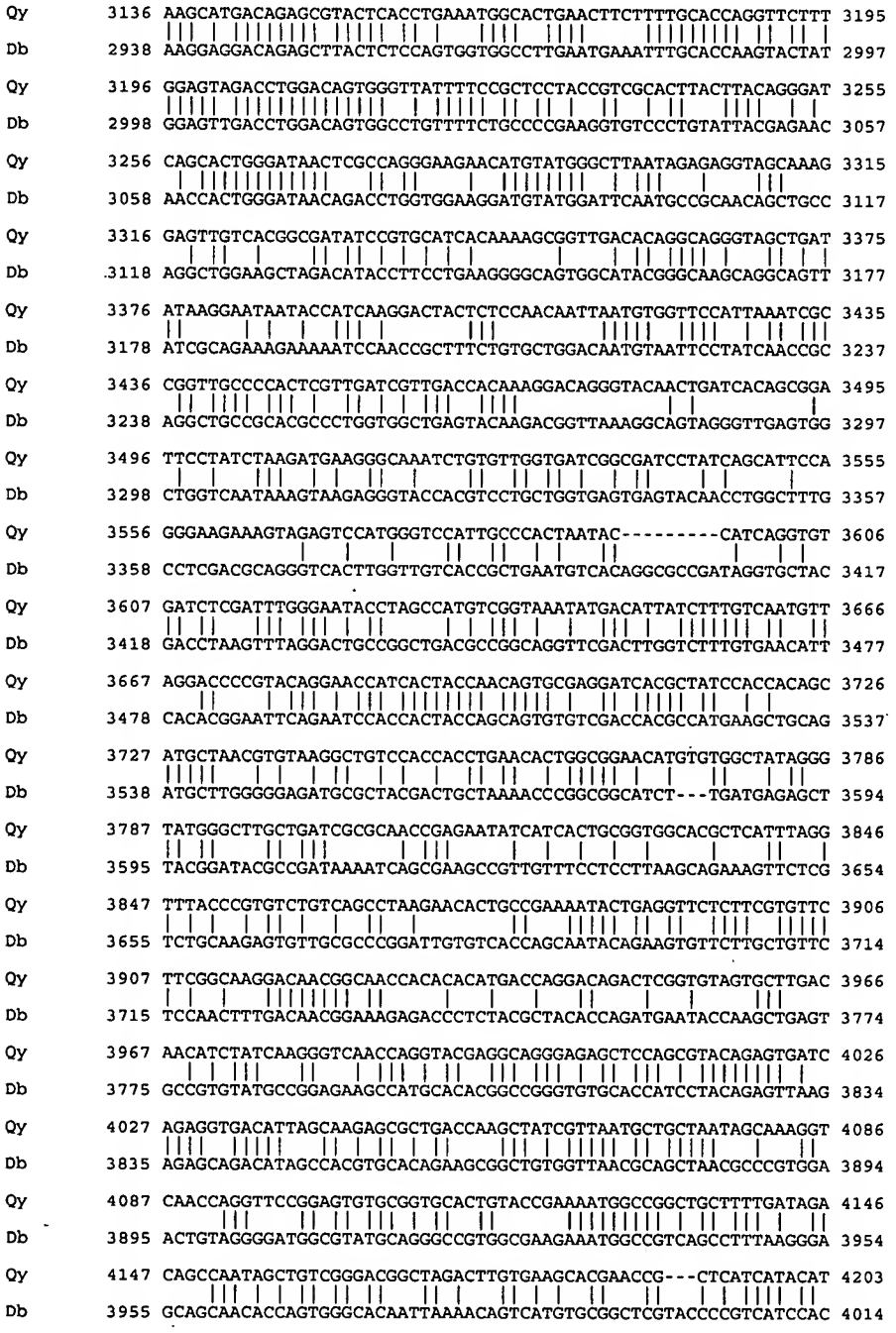
RESULT 7
US-09-554-337-4
; Sequence 4, Application US/09554337
; Patent No. 6475780

; GENERAL INFORMATION:
; APPLICANT: Parrington, Mark
; APPLICANT: Li, Xiaomao
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: ALPHAVIRUS VECTORS FOR PARAMYXOVIRUS VACCINES
; FILE REFERENCE: 1038-1042 MIS
; CURRENT APPLICATION NUMBER: US/09/554,337
; CURRENT FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/065,791
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: PCT/CA98/01064
; PRIOR FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 8100
; TYPE: DNA
; ORGANISM: respiratory syncytial virus
US-09-554-337-4

Query Match 29.5%; Score 1473.8; DB 4; Length 8100;
Best Local Similarity 57.9%; Pred. No. 0;
Matches 2779; Conservative 0; Mismatches 1982; Indels 36; Gaps 8;

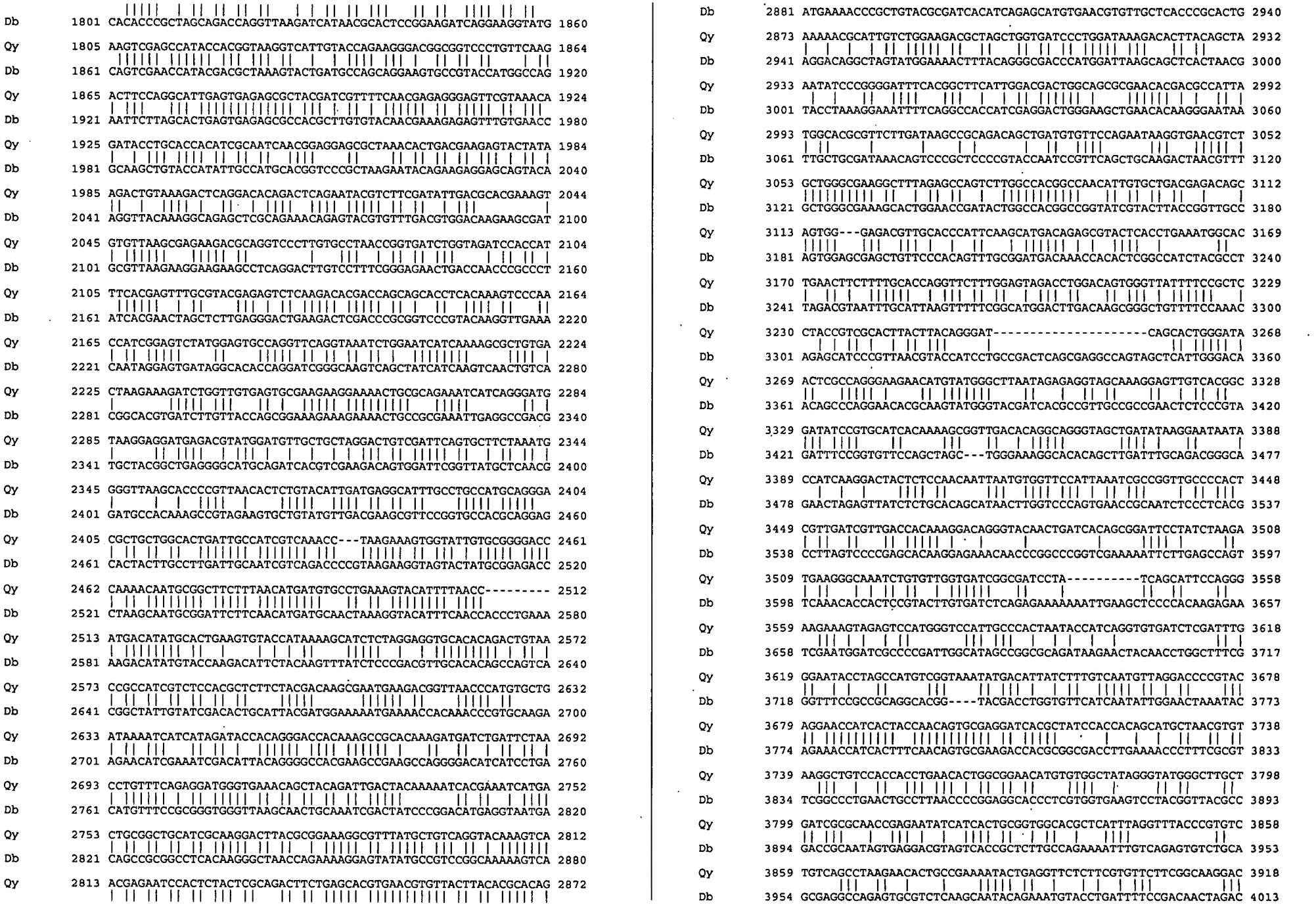



Db 721 GCCGTGACGTATCACGCGGAGGGATTCCTAGTGTGCAAGACCACAGACACTGTCAAAGGA 780
Qy 988 GAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCGATCAGATGACA 1047
Db 781 GAAAGAGTCTCATTCCCTGTATGCACCTACGTCCCCTCAACCATCTGTGATCAAATGACT 840
Qy 1048 GGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTATTGGTTGGGCTCAAC 1107
Db 841 GGCATACTAGCGACCGACGTCACACCGGAGGACGCACAGAAGTTGTTAGTGGGATTGAAT 900
Qy 1108 CAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAATGCAGAACTATCTA 1167
Db 901 CAGAGGATAGTTGTGAACGGAAGAACACAGCGAAACACTAACACGATGAAGAACTATCTG 960
Qy 1168 TTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATCGTGCCGACTTGGAC 1227
Db 961 CTTCCGATTGTGGCCGTCGCATTTAGCAAGTGGGCGAGGGAATACAAGGCAGACCTTGAT 1020
Qy 1228 GACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCTGCTGCTGGGCTTTC 1287
Db 1021 GATGAAAAACCTCTGGGTGTCCGAGAGAGGTCACTTACTTGCTGCTGCTTGTGGGCATTT 1080
Qy 1288 AAGACCCAGAAAATCACATCCATCTACAAGAAGCCTGGTACGCAAACAATTAAGAAAGTA 1347
Db 1081 AAAACGAGGAAGATGCACACCATGTACAAGAAACCAGACACCCAGACAATAGTGAAGGTG 1140
Qy 1348 CCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCACGGGCTCGATATGGGC 1407
Db 1141 CCTTCAGAGTTTAACTCGTTCGTCATCCCGAGCCTATGGTCTACAGGCCTCGCAATCCCA 1200
Qy 1408 TTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACTGTCAAACCCGCACCGGCTATTACA 1467
Db 1201 GTCAGATCACGCATTAAGATGCTTTTGGCCAAGAAGACCAAGCGAGA---GTTAATACCT 1257
Qy 1468 ATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAGAAGTGGCTGCAGCG 1527
Db 1258 GTTCTCGACGCGTCGTCAGCCAGGGATGCTGAACAAGAGGAGAAGGAGAGGTTGGAGGCC 1317
Qy 1528 GAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAAAAGAGACCGTAGAG 1587
Db 1318 GAGCTGACTAGAGAAGCCTTACCACCCCTCGTCCCCATCGCGCCGGCGGAGACGGGAGTC 1377
Qy 1588 GCAGAAGTAGACCTCATTATGCAAGAG------GCAGGAGCAGGTAGCGTGGAGACACCA 1641
Db 1378 GTCGACGTCGACGTTGAAGAACTAGAGTATCACGCAGGTGCAGGGGTCGTGGAAACACCT 1437
Qy 1642 CGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAGATTGGGTCTTACGCTATA 1701
Db 1438 CGCAGCGCGTTGAAAGTCACCGCACAGCCGAACGACGTACTACTAGGAAATTACGTAGTT 1497
Qy 1702 CTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGTATCCACCCATTGGCGGAA 1761
Db 1498 CTGTCCCCGCAGACCGTGCTCAAGAGCTCCAAGTTGGCCCCCGTGCACCCTCTAGCAGAG 1557
Qy 1762 CAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATACAAAGTCGAGCCATACCAC 1821
Db 1558 CAGGTGAAAATAATAACACATAACGGGAGGGCCGGCGGTTACCAGGTCGACGGATATGAC 1617
Qy 1822 GGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAAGACTTCCAGGCATTGAGT 1881
Db 1618 GGCAGGGTCCTACTACCATGTGGATCGGCCATTCCGGTCCCTGAGTTTCAAGCTTTGAGC 1677
Qy 1882 GAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAACAGATACCTGCACCACATC 1941
Db 1678 GAGAGCGCCACTATGGTGTACAACGAAAGGGAGTTCGTCAACAGGAAACTATACCATATT 1737
Qy 1942 GCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTATAAGACTGTAAAGACTCAG 2001
Db 1738 GCCGTTCACGGACCGTCGCTGAACACCGACGAGGAGAACTACGAGAAAGTCAGAGCTGAA 1797
Qy 2002 GACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAGTGTGTTAAGCGAGAAGAC 2061
Db 1798 AGAACTGACGCCGAGTACGTGTTCGACGTAGATAAAAAATGCTGCGTCAAGAGAGAGGAA 1857
Qy 2062 GCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCATTTCACGAGTTTGCGTAC 2121
Db 1858 GCGTCGGGTTTGGTGTTGGTGGGAGAGCTAACCAACCCCCCGTTCCATGAATTCGCCTAC 1917
Qy 2122 GAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCAACCATCGGAGTCTATGGA 2181
Db 1918 GAAGGGCTGAAGATCAGGCCGTCGGCACCATATAAGACTACAGTAGTAGGAGTCTTTGGG 1977
Qy 2182 GTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTGACTAAGAAAGATCTGGTT 2241
Db 1978 GTTCCGGGATCAGGCAAGTCTGCTATTATTAAGAGCCTCGTGACCAAACACGATCTGGTC 2037
Qy 2242 GTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGATGTAAGGAGGATGAGACGT 2301
Db 2038 ACCAGCGGCAAGAAGGAGAACTGCCAGGAAATAGTTAACGACGTGAAGAAGCACCGCGGG 2097
Qy 2302 ATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAATGGGGTTAAGCACCCCGTT 2361
Db 2098 AAGGGGACAAGTAGGGAAAACAGTGACTCCATCCTGCTAAACGGGTGTCGTCGTGCCGTG 2157
Qy 2362 AACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGGACGCTGCTGGCACTGATT 2421
Db 2158 GACATCCTATATGTGGACGAGGCTTTCGCTTGCCATTCCGGTACTCTGCTGGCCCTAATT 2217
Qy 2422 GCCATCGTCAAACCT---AAGAAAGTGGTATTGTGCGGGGACCCAAAACAATGCGGCTTC 2478
Db 2218 GCTCTTGTTAAACCTCGGAGCAAAGTGGTGTTATGCGGAGACCCCAAGCAATGCGGATTC 2277
Qy 2479 TTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGCACTGAAGTGTACCAT 2538
Db 2278 TTCAATATGATGCAGCTTAAGGTGAACTTCAACCACAACATCTGCACTGAAGTATGTCAT 2337
Qy 2539 AAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATCGTCTCCACGCTCTTCTAC 2598
Db 2338 AAAAGTATATCCAGACGTTGCACGCGTCCAGTCACGGCCATCGTCTACGTTGCACTAC 2397
Qy 2599 GACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATCATCATAGATACCACAGGG 2658
Db 2398 GGAGGCAAGATGCGCACGACCAACCCGTGCAACAAACCCATAATCATAGACACCACAGGA 2457
Qy 2659 ACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGAGGATGGGTGAAACAG 2718
Db 2458 CAGACCAAGCCCAAGCCAGGAGACATCGTGTTAACATGCTTCCGAGGCTGGGCAAAGCAG 2517
Qy 2719 CTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCTGCATCGCAAGGACTTACG 2778
Db 2518 CTGCAGTTGGACTACCGTGGACACGAAGTCATGACAGCAGCAGCATCTCAGGGCCTCACC 2577
Qy 2779 CGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCACTCTACTCGCAGACT 2838
Db 2578 CGCAAAGGGGTATACGCCGTAAGGCAGAAGGTGAATGAAAATCCCTTGTATGCCCCTGCG 2637
Qy 2839 TCTGAGCACGTGAACGTGTTACTTACACGCACAGAAAAACGCATTGTCTGGAAGACGCTA 2898
Db 2638 TCGGAGCACGTGAATGTACTGCTGACGCGCACTGAGGATAGGCTGGTGTGGAAAACGCTG 2697
Qy 2899 GCTGGTGATCCCTGGATAAAGACACTTACAGCTAAATATCCCGGGGATTTCACGGCTTCA 2958
Db 2698 GCCGGCGATCCCTGGATTAAGGTCCTATCAAACATTCCACAGGGTAACTTTACGGCCACA 2757
Qy 2959 TTGGACGACTGGCAGCGCGAACACGACGCCATTATGGCACGCGTTCTTGATAAGCCGCAG 3018
Db 2758 TTGGAAGAATGGCAAGAAGAACACGACAAAATAATGAAGGTGATTGAAGGACCGGCTGCG 2817
Qy 3019 ACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAGGCTTTAGAGCCAGTC 3078
Db 2818 CCTGTGGACGCGTTCCAGAACAAAGCGAACGTGTGTTGGGCGAAAAGCCTGGTGCCTGTC 2877
Qy 3079 TTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAGAC---GTTGCACCCATTC 3135
Db 2878 CTGGACACTGCCGGAATCAGATTGACAGCAGAGGAGTGGAGCACCATAATTACAGCATTT 2937



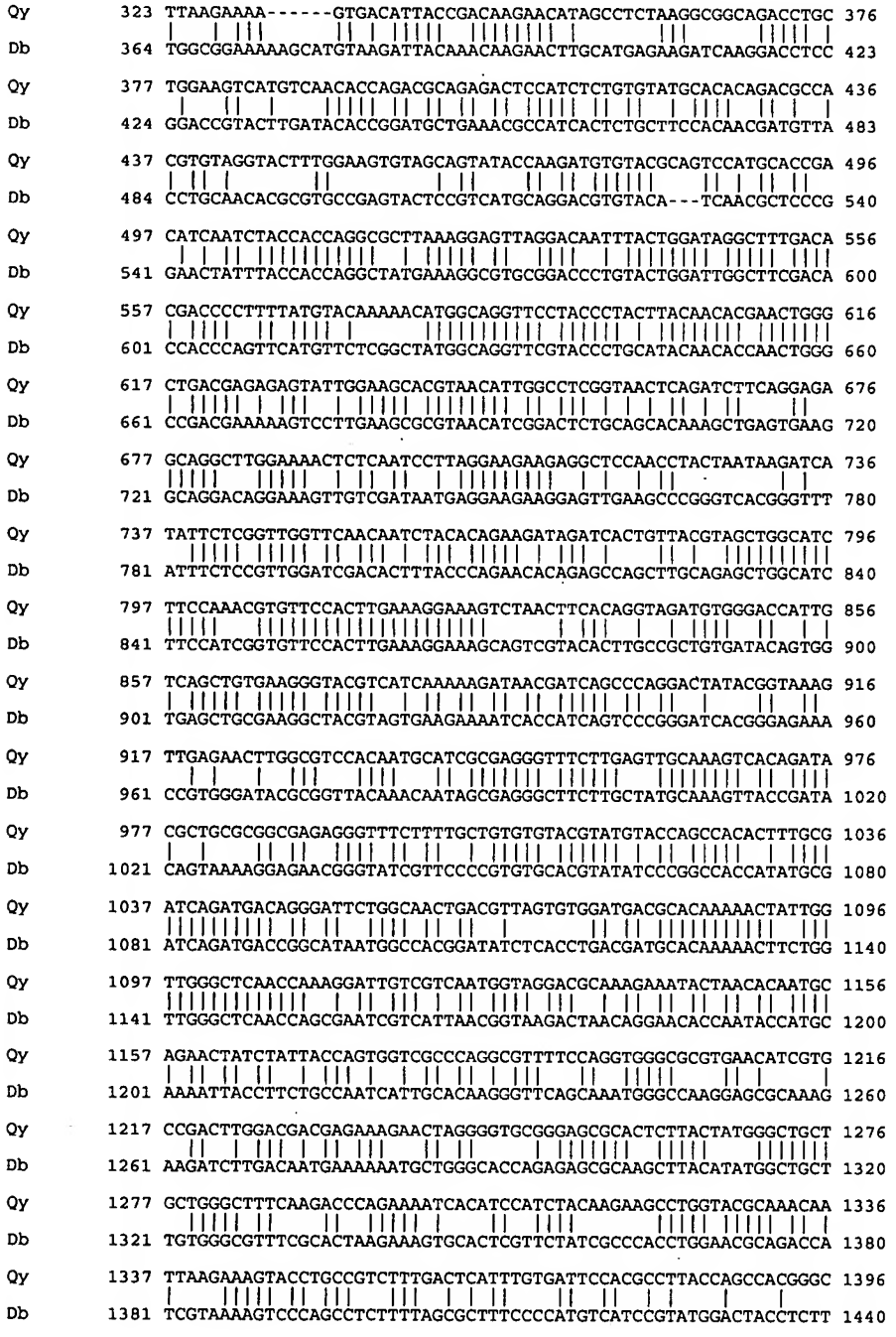
RESULT 8
US-08-446-932-1
; Sequence 1, Application US/08446932
; Patent No. 5639650
; GENERAL INFORMATION:
; APPLICANT: Johnston, Robert E.
; APPLICANT: Simpson, Dennis
; APPLICANT: Davis, Nancy L.
; TITLE OF INVENTION: cDNA Clone for South African
; TITLE OF INVENTION: Arbovirus No. 5639650 86
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth D. Sibley
; STREET: Post Office Drawer 34009
; CITY: Charlotte
; STATE: NC
; COUNTRY: USA
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

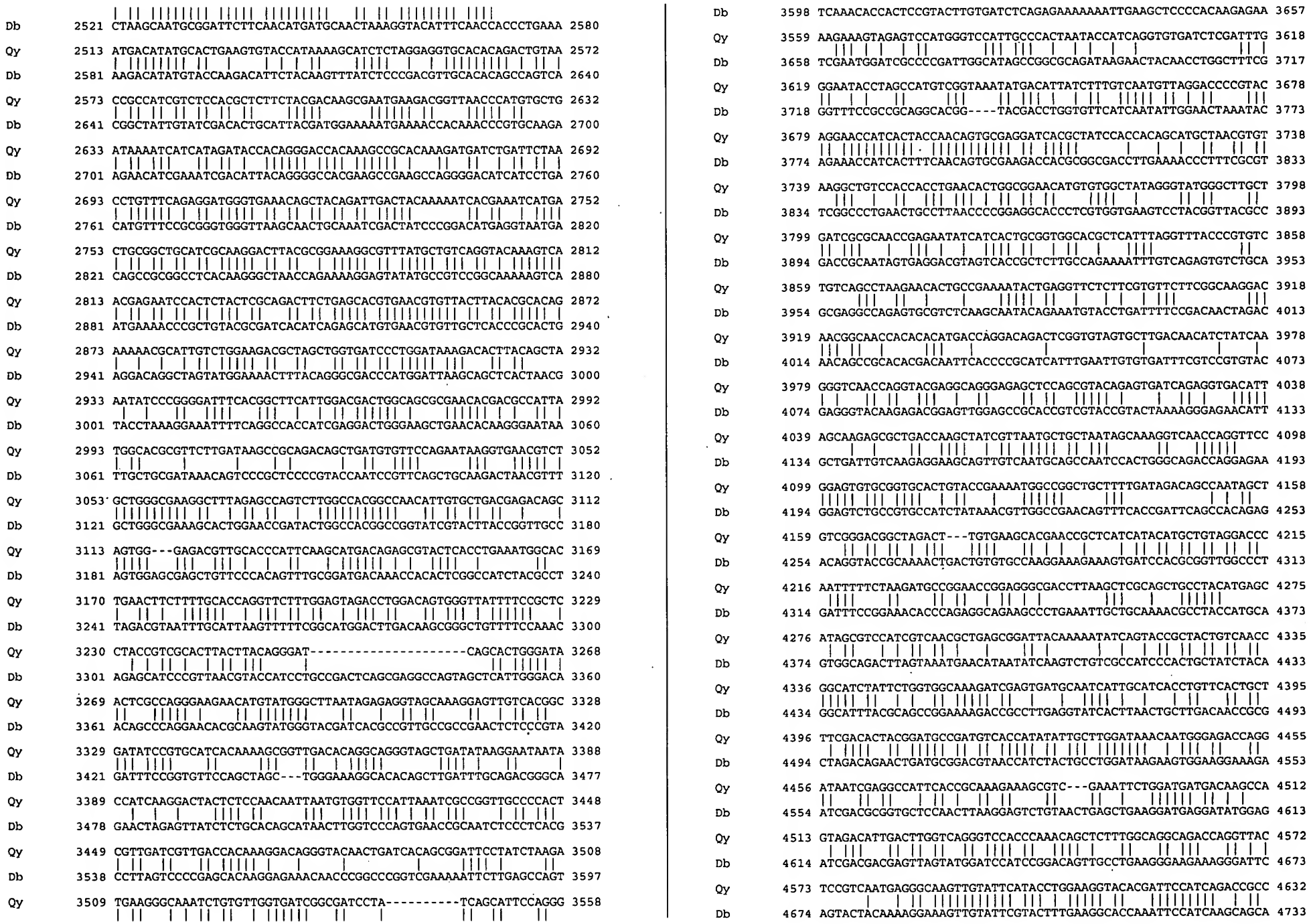
```
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/446,932
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sibley, Kenneth D.
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5470-118
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 881-3140
;      TELEFAX:  (919) 881-3175
;      TELEX:  575102
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11663 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
US-08-446-932-1

  Query Match             26.7%;  Score 1333.2;  DB 1;  Length 11663;
  Best Local Similarity   56.1%;  Pred. No. 0;
  Matches 2820;  Conservative    0;  Mismatches 2128;  Indels   80;  Gaps   13;

Qy         23 ATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGTCGTTACAGC 82
Db         64 AGAAGCCAGTAGTTAACGTAGACGTAGACCCTCAGAGTCCGTTTGTCGTGCAACTGCAAA 123

Qy         83 GGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACCATGCCAATG 142
Db        124 AGAGCTTCCCGCAATTTGAGGTAGTAGCACAGCAGGTCACTCCAAATGACCATGCTAATG 183

Qy        143 CCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACCGGGACCAAG 202
Db        184 CCAGAGCATTTTCGCATCTGGCCAGTAAACTAATCGAGCTGGAGGTTCCTACCACAGCGA 243

Qy        203 TTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACCGCTATCATT 262
Db        244 CGATTTTGGACATAGGCAGCGCACCGGCTCGTAGAATGTTTTCCGAGCACCAGTACCATT 303

Qy        263 GTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGGTATGCAGAAAGAC 322
Db        304 GCGTTTGCCCCATGCGTAGTCCAGAAGACCCGGACCGCATGATGAAATATGCCAGCAAAC 363

Qy        323 TTAAGAAAA------GTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACCTGC 376
Db        364 TGGCGGAAAAAGCATGTAAGATTACAAACAAGAACTTGCATGAGAAGATCAAGGACCTCC 423

Qy        377 TGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACGCCA 436
Db        424 GGACCGTACTTGATACACCGGATGCTGAAACGCCATCACTCTGCTTCCACAACGATGTTA 483

Qy        437 CGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCACCGA 496
Db        484 CCTGCAACACGCGTGCCGAGTACTCCGTCATGCAGGACGTGTACA---TCAACGCTCCCG 540

Qy        497 CATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTGACA 556
Db        541 GAACTATTTACCACCAGGCTATGAAAGGCGTGCGGACCCTGTACTGGATTGGCTTCGACA 600

Qy        557 CGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACTGGG 616
Db        601 CCACCCAGTTCATGTTCTCGGCTATGGCAGGTTCGTACCCTGCATACAACACCAACTGGG 660

Qy        617 CTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGGAGA 676
Db        661 CCGACGAAAAAGTCCTTGAAGCGCGTAACATCGGACTCTGCAGCACAAAGCTGAGTGAAG 720

Qy        677 GCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGATCA 736
Db        721 GCAGGACAGGAAAGTTGTCGATAATGAGGAAGAAGGAGTTGAAGCCCGGGTCACGGGTTT 780

Qy        737 TATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGCATC 796
Db        781 ATTTCTCCGTTGGATCGACACTTTACCCAGAACACAGAGCCAGCTTGCAGAGCTGGCATC 840

Qy        797 TTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTG 856
Db        841 TTCCATCGGTGTTCCACTTGAAAGGAAAGCAGTCGTACACTTGCCGCTGTGATACAGTGG 900

Qy        857 TCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAG 916
Db        901 TGAGCTGCGAAGGCTACGTAGTGAAGAAAATCACCATCAGTCCCGGGATCACGGGAGAAA 960

Qy        917 TTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATA 976
Db        961 CCGTGGGATACGCGGTTACAAACAATAGCGAGGGCTTCTTGCTATGCAAAGTTACCGATA 1020

Qy        977 CGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCG 1036
Db       1021 CAGTAAAAGGAGAACGGGTATCGTTCCCCGTGTGCACGTATATCCCGGCCACCATATGCG 1080

Qy       1037 ATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTATTGG 1096
Db       1081 ATCAGATGACCGGCATAATGGCCACGGATATCTCACCTGACGATGCACAAAAACTTCTGG 1140

Qy       1097 TTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAATGC 1156
Db       1141 TTGGGCTCAACCAGCGAATCGTCATTAACGGTAAGACTAACAGGAACACCAATACCATGC 1200

Qy       1157 AGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATCGTG 1216
Db       1201 AAAATTACCTTCTGCCAATCATTGCACAAGGGTTCAGCAAATGGGCCAAGGAGCGCAAAG 1260

Qy       1217 CCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCTGCT 1276
Db       1261 AAGATCTTGACAATGAAAAAATGCTGGGCACCAGAGAGCGCAAGCTTACATATGGCTGCT 1320

Qy       1277 GCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCTGGTACGCAAACAA 1336
Db       1321 TGTGGGCGTTTCGCACTAAGAAAGTGCACTCGTTCTATCGCCCACCTGGAACGCAGACCA 1380

Qy       1337 TTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCACGGGC 1396
Db       1381 TCGTAAAAGTCCCAGCCTCTTTTAGCGCTTTCCCCATGTCATCCGTATGGACTACCTCTT 1440

Qy       1397 TCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCA---ACTGTCAAACCCG 1453
Db       1441 TGCCCATGTCGCTGAGGCAGAAGATGAAATTGGCATTACAACCAAAGAAGGAGGAAAAAC 1500

Qy       1454 CACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAG 1513
Db       1501 TGCTGCAAGTCCCGGAGGAATTAGTTATGGAGGCCAAGGCTGCTTTCGAGGATGCTCAGG 1560

Qy       1514 AAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAA 1573
Db       1561 AGGAATCCAGAGCGGAGAAGCTCCGAGAAGCACTCCCACCATTAGTGGCAGACAAAGGTA 1620

Qy       1574 AAGAGACCGTAGAGGCAG---------AAGTAGACCTCATTATGCAAGAGGCAGGAGCAG 1624
Db       1621 TCGAGGCAGCTGCGGAAGTTGTCTGCGAAGTGGAGGGGCTCCAGGCGGACACCGGAGCAG 1680

Qy       1625 GTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAGA 1684
Db       1681 CACTCGTCGAAACCCCGCGCGGTCATGTAAGGATAATACCTCAAGCAAATGACCGTATGA 1740

Qy       1685 TTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGTA 1744
Db       1741 TCGGACAGTATATCGTTGTCTCGCCGATCTCTGTGCTGAAGAACGCTAAACTCGCACCAG 1800

Qy       1745 TCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATACA 1804
```



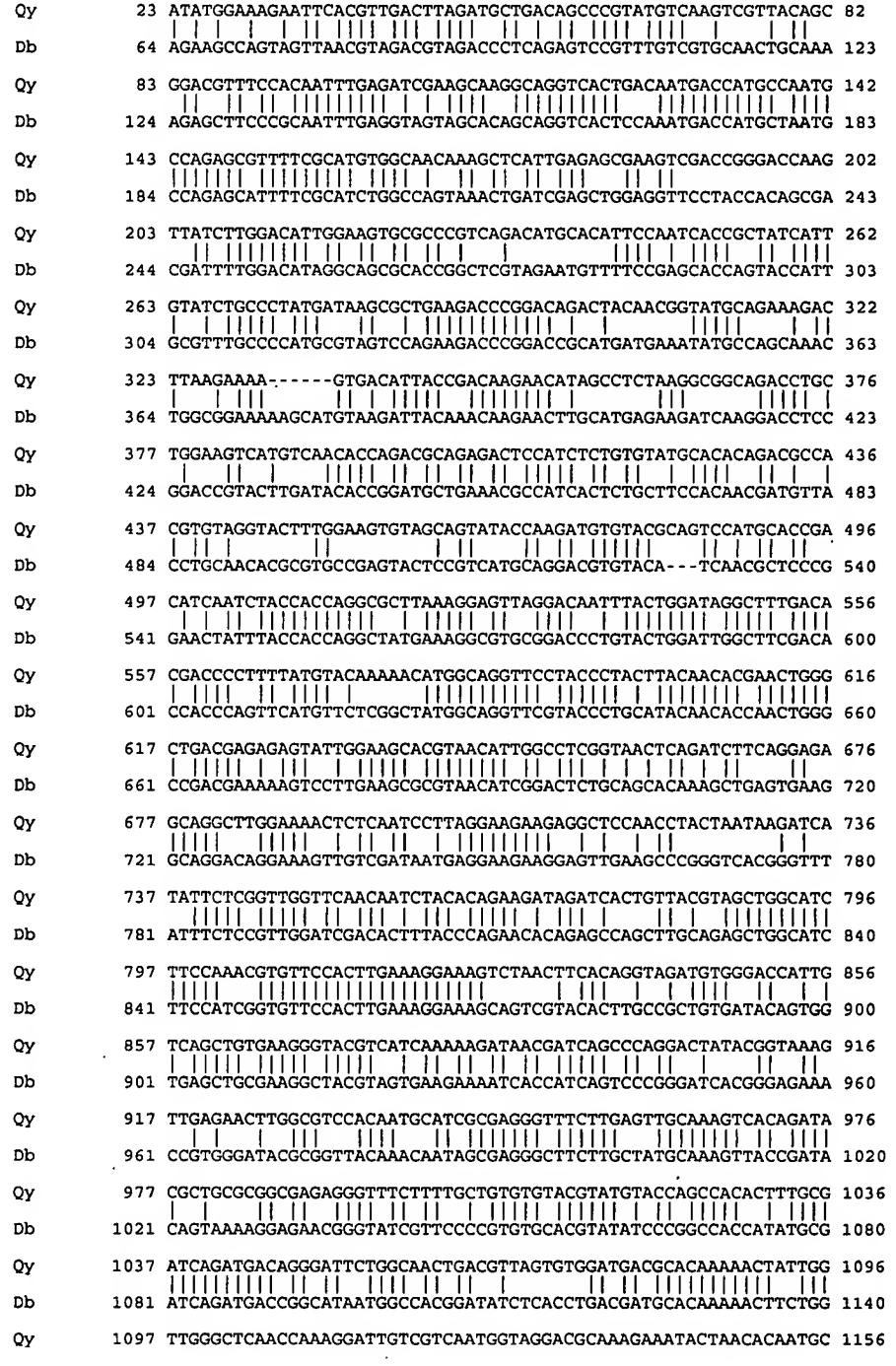
```
Qy       3919 AACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTGCTTGACAACATCTATCAA 3978
Db       4014 AACAGCCGCACACGACAATTCACCCCGCATCATTTGAATTGTGTGATTTCGTCCGTGTAC 4073

Qy       3979 GGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGTACAGAGTGATCAGAGGTGACATT 4038
Db       4074 GAGGGTACAAGAGACGGAGTTGGAGCCGCACCGTCGTACCGTACTAAAAGGGAGAACATT 4133

Qy       4039 AGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGCAAAGGTCAACCAGGTTCC 4098
Db       4134 GCTGATTGTCAAGAGGAAGCAGTTGTCAATGCAGCCAATCCACTGGGCAGACCAGGAGAA 4193

Qy       4099 GGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTTGATAGACAGCCAATAGCT 4158
Db       4194 GGAGTCTGCCGTGCCATCTATAAACGTTGGCCGAACAGTTTCACCGATTCAGCCACAGAG 4253

Qy       4159 GTCGGGACGGCTAGACT---TGTGAAGCACGAACCGCTCATCATACATGCTGTAGGACCC 4215
Db       4254 ACAGGTACCGCAAAACTGACTGTGTGCCAAGGAAAGAAAGTGATCCACGCGGTTGGCCCT 4313

Qy       4216 AATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTAAGCTCGCAGCTGCCTACATGAGC 4275
Db       4314 GATTTCCGGAAACACCCAGAGGCAGAAGCCCTGAAATTGCTGCAAAACGCCTACCATGCA 4373

Qy       4276 ATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTACCGCTACTGTCAACC 4335
Db       4374 GTGGCAGACTTAGTAAATGAACATAATATCAAGTCTGTCGCCATCCCACTGCTATCTACA 4433

Qy       4336 GGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCATCACCTGTTCACTGCT 4395
Db       4434 GGCATTTACGCAGCCGGAAAAGACCGCCTTGAGGTATCACTTAACTGCTTGACAACCGCG 4493

Qy       4396 TTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAACAATGGGAGACCAGG 4455
Db       4494 CTAGACAGAACTGATGCGGACGTAACCATCTACTGCCTGGATAAGAAGTGGAAGGAAAGA 4553

Qy       4456 ATAATCGAGGCCATTCACCGCAAAGAAAGCGTC---GAAATTCTGGATGATGACAAGCCA 4512
Db       4554 ATCGACGCGGTGCTCCAACTTAAGGAGTCTGTAACTGAGCTGAAGGATGAGGATATGGAG 4613

Qy       4513 GTAGACATTGACTTGGTCAGGGTCCACCCAAACAGCTCTTTGGCAGGCAGACCAGGTTAC 4572
Db       4614 ATCGACGACGAGTTAGTATGGATCCATCCGGACAGTTGCCTGAAGGGAAGAAAGGGATTC 4673

Qy       4573 TCCGTCAATGAGGGCAAGTTGTATTCATACCTGGAAGGTACACGATTCCATCAGACCGCC 4632
Db       4674 AGTACTACAAAAGGAAAGTTGTATTCGTACTTTGAAGGCACCAAATTCCATCAAGCAGCA 4733

Qy       4633 AAGGACATTGCCGAAATCCATGCAATGTGGCCCAACAAATCTGAGGCTAATGAGCAGATT 4692
Db       4734 AAAGATATGGCGGAGATAAAGGTCCTGTTCCCAAATGACCAGGAAAGCAACGAACAACTG 4793

Qy       4693 TGCTTGTACATCCTGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGCCCAGTAGAGGAG 4752
Db       4794 TGTGCCTACATATTGGGGGAGACCATGGAAGCAATCCGCGAAAAATGCCCGGTCGACCAC 4853

Qy       4753 TCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTACGCTATGACGGCT 4812
Db       4854 AACCCGTCGTCTAGCCCGCCAAAAACGCTGCCGTGCCTCTGTATGTATGCCATGACGCCA 4913

Qy       4813 GAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTATGCTCATCATTC 4872
Db       4914 GAAAGGGTCCACAGACTCAGAAGCAATAACGTCAAAGAAGTTACAGTATGCTCCTCCACC 4973

Qy       4873 CTGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGCAAACCAGTCCTG 4932
Db       4974 CCCCTTCCAAAGTACAAAATCAAGAATGTTCAGAAGGTTCAGTGCACAAAAGTAGTCCTG 5033

Qy       4933 TTTTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGCGGAA 4980
Db       5034 TTTAACCCGCATACCCCCGCATTCGTTCCCGCCCGTAAGTACATAGAA 5081
```

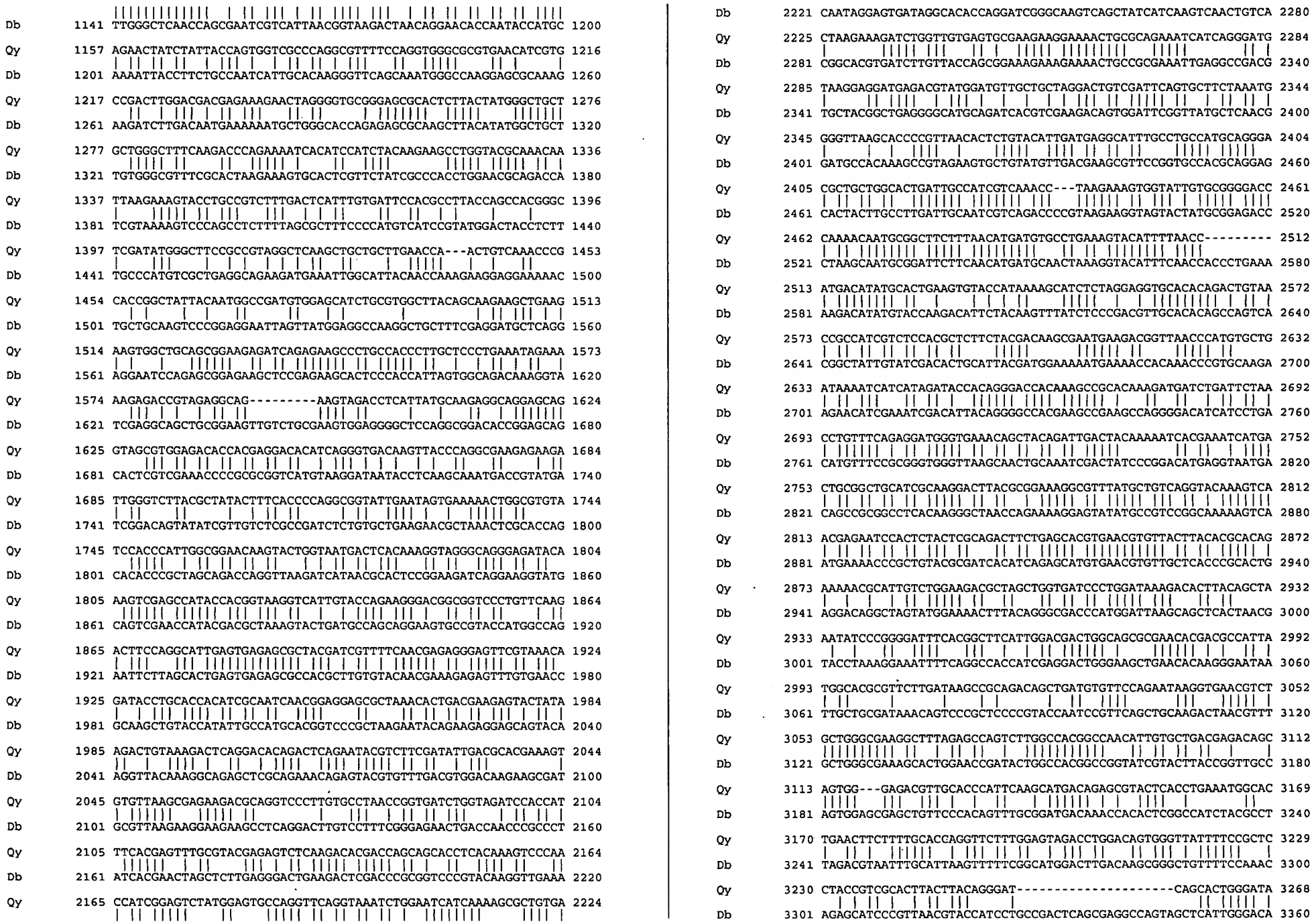
```
RESULT 9
US-08-801-263A-1
; Sequence 1, Application US/08801263A
; Patent No. 5811407
;  GENERAL INFORMATION:
;    APPLICANT:  Johnston, Robert E.
;    APPLICANT:  Davis, Nancy L.
;    APPLICANT:  Simpson, Dennis A.
;    TITLE OF INVENTION:  System for the In Vivo Delivery and
;    TITLE OF INVENTION:  Expression of Heterologous Genes in the Bone Marrow
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson, P.A.
;      STREET:  1211 East Morehead Street
;      CITY:  Charlotte
;      STATE:  No. 5811407th Carolina
;      COUNTRY:  USA
;      ZIP:  28234
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/801,263A
;      FILING DATE:  19-FEB-1997
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sibley, Kenneth D.
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5470-147
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-420-2200
;      TELEFAX:  919-881-3175
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11663 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  60..7559
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  7608..11342
US-08-801-263A-1

  Query Match             26.7%;  Score 1333.2;  DB 1;  Length 11663;
  Best Local Similarity   56.1%;  Pred. No. 0;
  Matches 2820;  Conservative    0;  Mismatches 2128;  Indels   80;  Gaps   13;

Qy         23 ATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGTCGTTACAGC 82
Db         64 AGAAGCCAGTAGTTAACGTAGACGTAGACCCTCAGAGTCCGTTTGTCGTGCAACTGCAAA 123

Qy         83 GGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACCATGCCAATG 142
Db        124 AGAGCTTCCCGCAATTTGAGGTAGTAGCACAGCAGGTCACTCCAAATGACCATGCTAATG 183

Qy        143 CCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACCGGGACCAAG 202
Db        184 CCAGAGCATTTTCGCATCTGGCCAGTAAACTAATCGAGCTGGAGGTTCCTACCACAGCGA 243

Qy        203 TTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACCGCTATCATT 262
Db        244 CGATTTTGGACATAGGCAGCGCACCGGCTCGTAGAATGTTTTCCGAGCACCAGTACCATT 303

Qy        263 GTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGGTATGCAGAAAGAC 322
Db        304 GCGTTTGCCCCATGCGTAGTCCAGAAGACCCGGACCGCATGATGAAATATGCCAGCAAAC 363
```

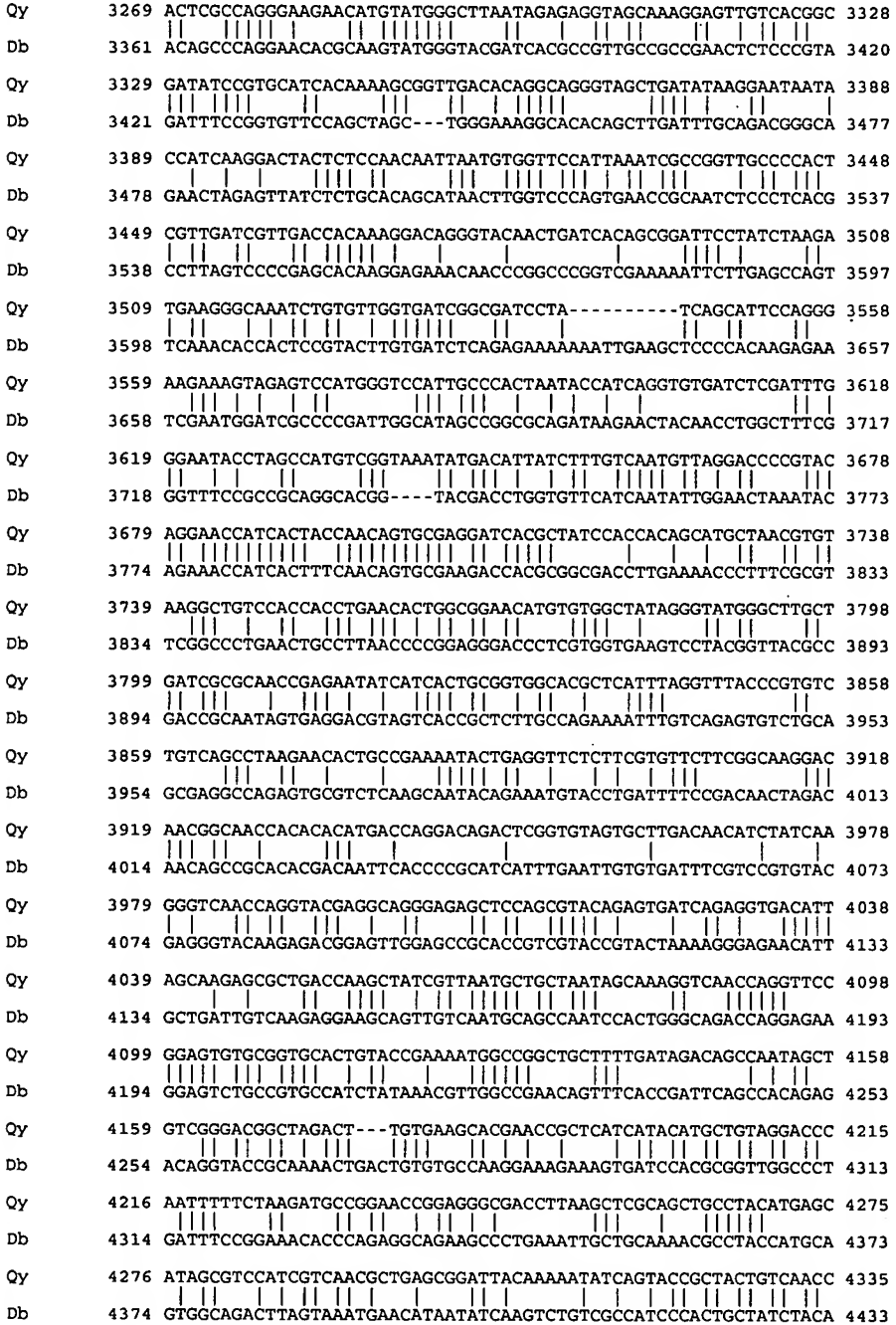

Qy 323 TTAAGAAAA------GTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACCTGC 376
Db 364 TGGCGGAAAAAGCATGTAAGATTACAAACAAGAACTTGCATGAGAAGATCAAGGACCTCC 423
Qy 377 TGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACGCCA 436
Db 424 GGACCGTACTTGATACACCGGATGCTGAAACGCCATCACTCTGCTTCCACAACGATGTTA 483
Qy 437 CGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCACCGA 496
Db 484 CCTGCAACACGCGTGCCGAGTACTCCGTCATGCAGGACGTGTACA---TCAACGCTCCCG 540
Qy 497 CATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTGACA 556
Db 541 GAACTATTTACCACCAGGCTATGAAAGGCGTGCGGACCCTGTACTGGATTGGCTTCGACA 600
Qy 557 CGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACTGGG 616
Db 601 CCACCCAGTTCATGTTCTCGGCTATGGCAGGTTCGTACCCTGCATACAACACCAACTGGG 660
Qy 617 CTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGGAGA 676
Db 661 CCGACGAAAAAGTCCTTGAAGCGCGTAACATCGGACTCTGCAGCACAAAGCTGAGTGAAG 720
Qy 677 GCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGATCA 736
Db 721 GCAGGACAGGAAAGTTGTCGATAATGAGGAAGAAGGAGTTGAAGCCCGGGTCACGGGTTT 780
Qy 737 TATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGCATC 796
Db 781 ATTTCTCCGTTGGATCGACACTTTACCCAGAACACAGAGCCAGCTTGCAGAGCTGGCATC 840
Qy 797 TTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTG 856
Db 841 TTCCATCGGTGTTCCACTTGAAAGGAAAGCAGTCGTACACTTGCCGCTGTGATACAGTGG 900
Qy 857 TCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAG 916
Db 901 TGAGCTGCGAAGGCTACGTAGTGAAGAAAATCACCATCAGTCCCGGGATCACGGGAGAAA 960
Qy 917 TTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATA 976
Db 961 CCGTGGGATACGCGGTTACAAACAATAGCGAGGGCTTCTTGCTATGCAAAGTTACCGATA 1020
Qy 977 CGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCG 1036
Db 1021 CAGTAAAAGGAGAACGGGTATCGTTCCCCGTGTGCACGTATATCCCGGCCACCATATGCG 1080
Qy 1037 ATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTATTGG 1096
Db 1081 ATCAGATGACCGGCATAATGGCCACGGATATCTCACCTGACGATGCACAAAAACTTCTGG 1140
Qy 1097 TTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAATGC 1156
Db 1141 TTGGGCTCAACCAGCGAATCGTCATTAACGGTAAGACTAACAGGAACACCAATACCATGC 1200
Qy 1157 AGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATCGTG 1216
Db 1201 AAAATTACCTTCTGCCAATCATTGCACAAGGGTTCAGCAAATGGGCCAAGGAGCGCAAAG 1260
Qy 1217 CCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCTGCT 1276
Db 1261 AAGATCTTGACAATGAAAAAATGCTGGGCACCAGAGAGCGCAAGCTTACATATGGCTGCT 1320
Qy 1277 GCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCTGGTACGCAAACAA 1336
Db 1321 TGTGGGCGTTTCGCACTAAGAAAGTGCACTCGTTCTATCGCCCACCTGGAACGCAGACCA 1380
Qy 1337 TTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCACGGGC 1396
Db 1381 TCGTAAAAGTCCCAGCCTCTTTTAGCGCTTTCCCCATGTCATCCGTATGGACTACCTCTT 1440
Qy 1397 TCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCA---ACTGTCAAACCCG 1453
Db 1441 TGCCCATGTCGCTGAGGCAGAAGATGAAATTGGCATTACAACCAAAGAAGGAGGAAAAAC 1500
Qy 1454 CACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAG 1513
Db 1501 TGCTGCAAGTCCCGGAGGAATTAGTTATGGAGGCCAAGGCTGCTTTCGAGGATGCTCAGG 1560
Qy 1514 AAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAA 1573
Db 1561 AGGAATCCAGAGCGGAGAAGCTCCGAGAAGCACTCCCACCATTAGTGGCAGACAAAGGTA 1620
Qy 1574 AAGAGACCGTAGAGGCAG---------AAGTAGACCTCATTATGCAAGAGGCAGGAGCAG 1624
Db 1621 TCGAGGCAGCTGCGGAAGTTGTCTGCGAAGTGGAGGGGCTCCAGGCGGACACCGGAGCAG 1680
Qy 1625 GTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAGA 1684
Db 1681 CACTCGTCGAAACCCCGCGCGGTCATGTAAGGATAATACCTCAAGCAAATGACCGTATGA 1740
Qy 1685 TTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGTA 1744
Db 1741 TCGGACAGTATATCGTTGTCTCGCCGATCTCTGTGCTGAAGAACGCTAAACTCGCACCAG 1800
Qy 1745 TCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATACA 1804
Db 1801 CACACCCGCTAGCAGACCAGGTTAAGATCATAACGCACTCCGGAAGATCAGGAAGGTATG 1860
Qy 1805 AAGTCGAGCCATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAAG 1864
Db 1861 CAGTCGAACCATACGACGCTAAAGTACTGATGCCAGCAGGAAGTGCCGTACCATGGCCAG 1920
Qy 1865 ACTTCCAGGCATTGAGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAACA 1924
Db 1921 AATTCTTAGCACTGAGTGAGAGCGCCACGCTTGTGTACAACGAAAGAGAGTTTGTGAACC 1980
Qy 1925 GATACCTGCACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTATA 1984
Db 1981 GCAAGCTGTACCATATTGCCATGCACGGTCCCGCTAAGAATACAGAAGAGGAGCAGTACA 2040
Qy 1985 AGACTGTAAAGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAGT 2044
Db 2041 AGGTTACAAAGGCAGAGCTCGCAGAAACAGAGTACGTGTTTGACGTGGACAAGAAGCGAT 2100
Qy 2045 GTGTTAAGCGAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCAT 2104
Db 2101 GCGTTAAGAAGGAAGAAGCCTCAGGACTTGTCCTTTCGGGAGAACTGACCAACCCGCCCT 2160
Qy 2105 TTCACGAGTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCAA 2164
Db 2161 ATCACGAACTAGCTCTTGAGGGACTGAAGACTCGACCCGCGGTCCCGTACAAGGTTGAAA 2220
Qy 2165 CCATCGGAGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTGA 2224
Db 2221 CAATAGGAGTGATAGGCACACCAGGATCGGGCAAGTCAGCTATCATCAAGTCAACTGTCA 2280
Qy 2225 CTAAGAAAGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGATG 2284
Db 2281 CGGCACGTGATCTTGTTACCAGCGGAAAGAAAGAAAACTGCCGCGAAATTGAGGCCGACG 2340
Qy 2285 TAAGGAGGATGAGACGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAATG 2344
Db 2341 TGCTACGGCTGAGGGGCATGCAGATCACGTCGAAGACAGTGGATTCGGTTATGCTCAACG 2400
Qy 2345 GGGTTAAGCACCCCGTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGGA 2404
Db 2401 GATGCCACAAAGCCGTAGAAGTGCTGTATGTTGACGAAGCGTTCCGGTGCCACGCAGGAG 2460
Qy 2405 CGCTGCTGGCACTGATTGCCATCGTCAAACC---TAAGAAAGTGGTATTGTGCGGGGACC 2461
Db 2461 CACTACTTGCCTTGATTGCAATCGTCAGACCCCGTAAGAAGGTAGTACTATGCGGAGACC 2520
Qy 2462 CAAAACAATGCGGCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAACC--------- 2512

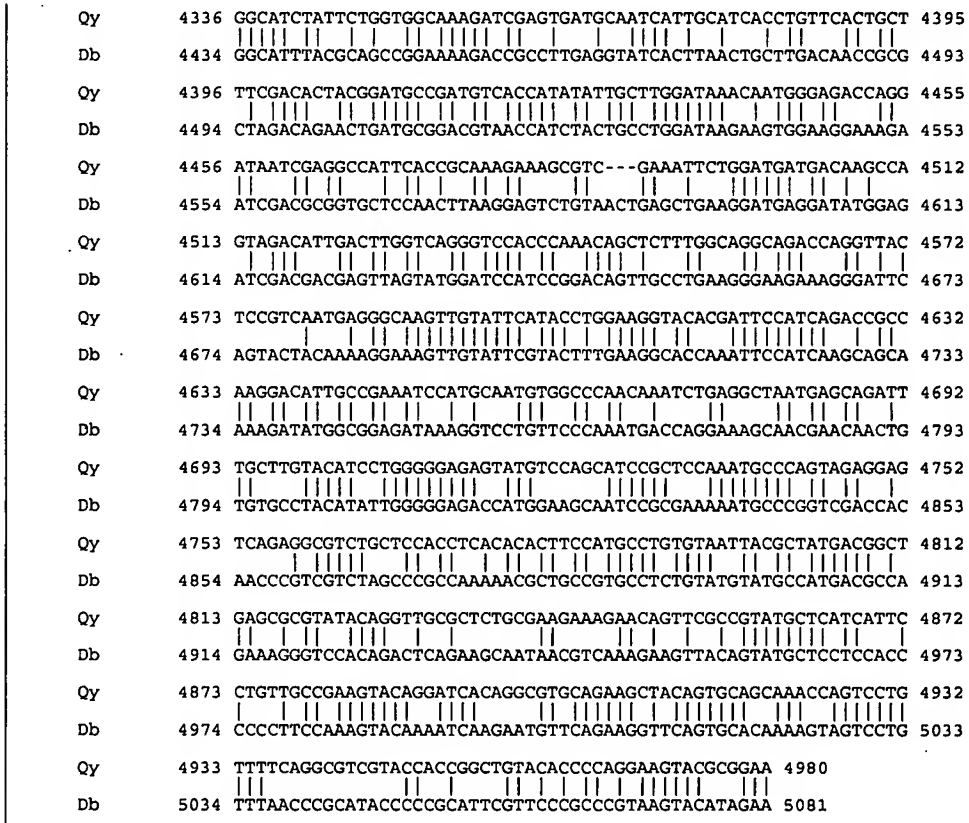


```
RESULT 10
US-08-801-263A-7
; Sequence 7, Application US/08801263A
; Patent No. 5811407
;  GENERAL INFORMATION:
;    APPLICANT:  Johnston, Robert E.
;    APPLICANT:  Davis, Nancy L.
;    APPLICANT:  Simpson, Dennis A.
;    TITLE OF INVENTION:  System for the In Vivo Delivery and
;    TITLE OF INVENTION:  Expression of Heterologous Genes in the Bone Marrow
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson, P.A.
;      STREET:  1211 East Morehead Street
;      CITY:  Charlotte
;      STATE:  No. 5811407th Carolina
;      COUNTRY:  USA
;      ZIP:  28234
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/801,263A
;      FILING DATE:  19-FEB-1997
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sibley, Kenneth D.
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5470-147
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-420-2200
;      TELEFAX:  919-881-3175
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11663 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-801-263A-7

  Query Match             26.7%;  Score 1333.2;  DB 1;  Length 11663;
  Best Local Similarity   56.1%;  Pred. No. 0;
  Matches 2820;  Conservative    0;  Mismatches 2128;  Indels   80;  Gaps   13;
```





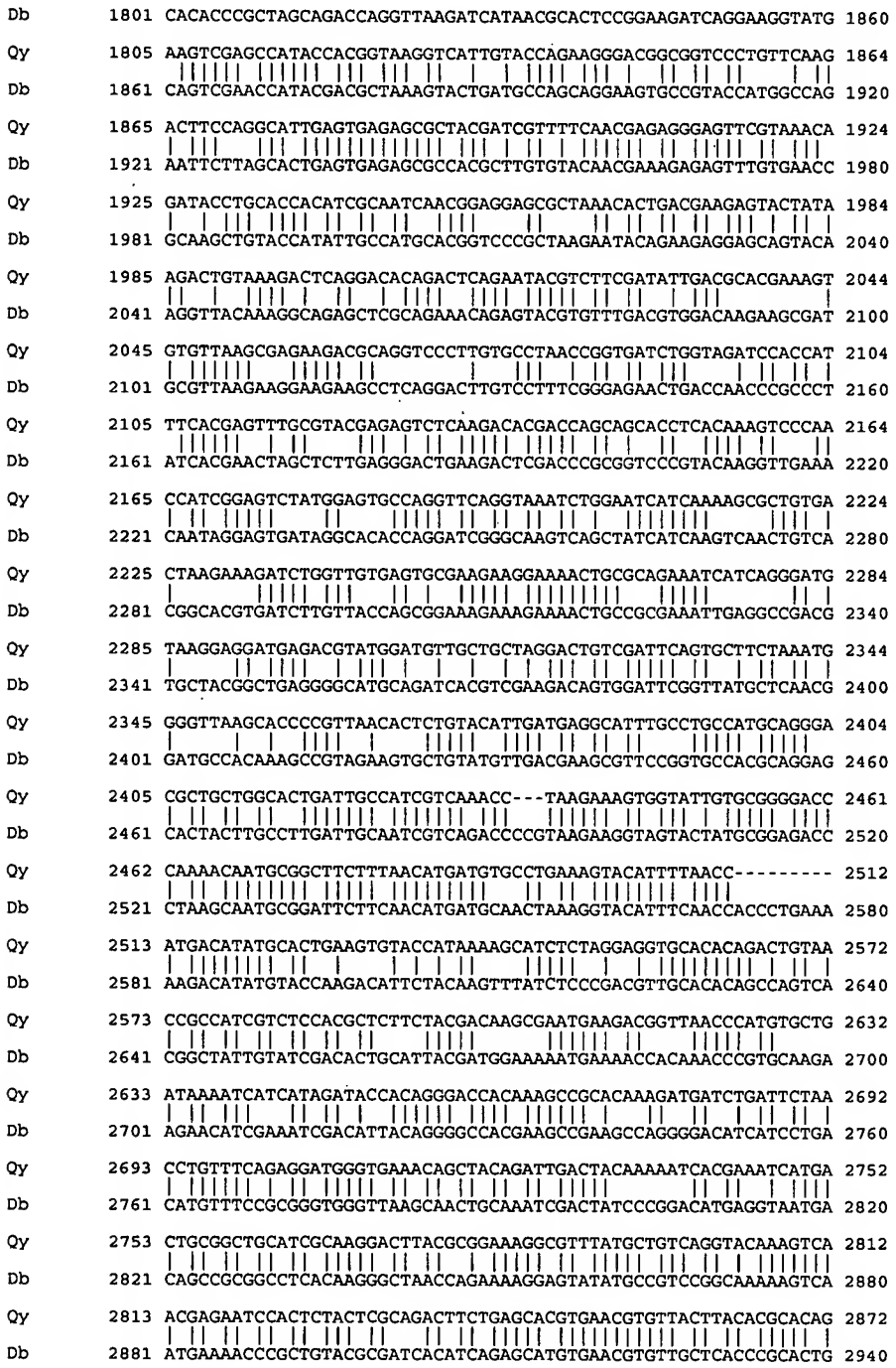


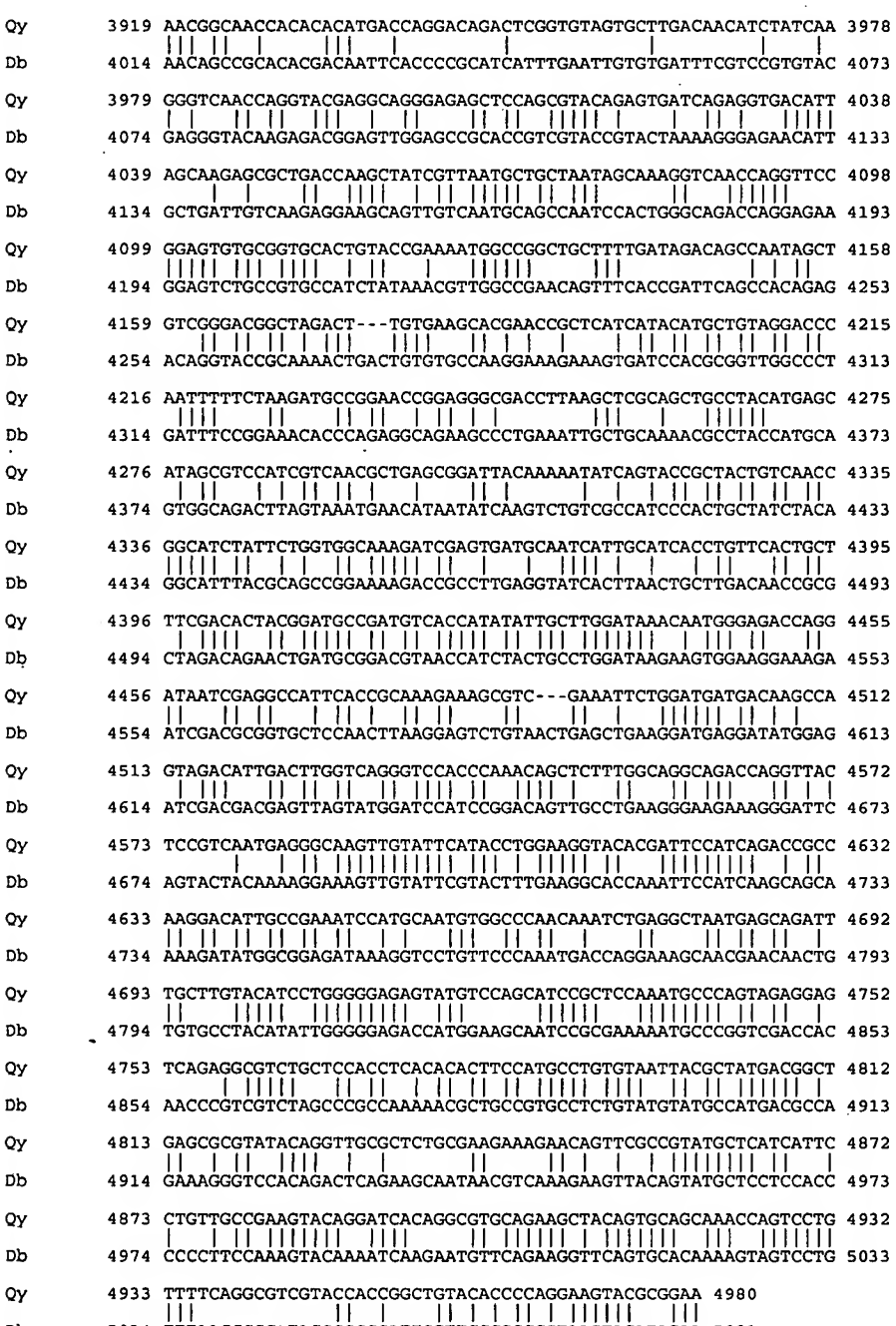


RESULT 11
US-09-102-248-1
; Sequence 1, Application US/09102248
; Patent No. 6008035
; GENERAL INFORMATION:
; APPLICANT: Johnston, Robert E.
; APPLICANT: Davis, Nancy L.
; APPLICANT: Simpson, Dennis A.
; TITLE OF INVENTION: System for the In Vivo Delivery and
; TITLE OF INVENTION: Expression of Heterologous Genes in the Bone Marrow
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson, P.A.
; STREET: 1211 East Morehead Street
; CITY: Charlotte
; STATE: No. 6008035th Carolina
; COUNTRY: USA
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/102,248
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

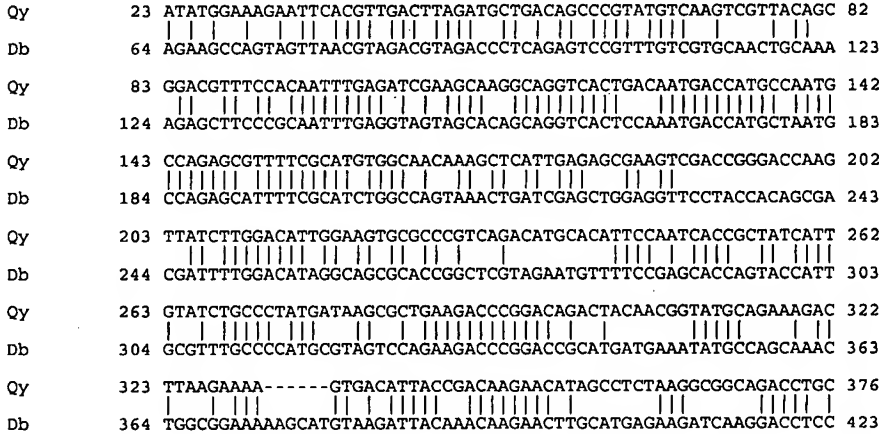
```
;     APPLICATION NUMBER:  US 08/801,263
;       FILING DATE:  19-FEB-1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Sibley, Kenneth D.
;       REGISTRATION NUMBER:  31,665
;       REFERENCE/DOCKET NUMBER:  5470-147
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  919-420-2200
;       TELEFAX:  919-881-3175
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  11663 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  60..7559
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  7608..11342
US-09-102-248-1

  Query Match           26.7%;  Score 1333.2;  DB 3;  Length 11663;
  Best Local Similarity  56.1%;  Pred. No. 0;
  Matches 2820;  Conservative    0;  Mismatches 2128;  Indels   80;  Gaps   13;

Qy         23 ATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGTCGTTACAGC 82
              | | |  |  | || |||| ||| ||||  || | || |||| ||||  |    | ||
Db         64 AGAAGCCAGTAGTTAACGTAGACGTAGACCCTCAGAGTCCGTTTGTCGTGCAACTGCAAA 123

Qy         83 GGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACCATGCCAATG 142
               ||  || || ||||||||| | | ||||  ||||||||||   ||||||||||| ||||
Db        124 AGAGCTTCCCGCAATTTGAGGTAGTAGCACAGCAGGTCACTCCAAATGACCATGCTAATG 183

Qy        143 CCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACCGGGACCAAG 202
              ||||||| ||||||||| |||| |  || || || |||   || ||
Db        184 CCAGAGCATTTTCGCATCTGGCCAGTAAACTAATCGAGCTGGAGGTTCCTACCACAGCGA 243

Qy        203 TTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACCGCTATCATT 262
               || |||||||| || || || || |   |          |||| | ||||  || ||||
Db        244 CGATTTTGGACATAGGCAGCGCACCGGCTCGTAGAATGTTTTCCGAGCACCAGTACCATT 303

Qy        263 GTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGGTATGCAGAAAGAC 322
              |  | ||||| |||   ||  | ||||||||||||| | |     |||||    | ||
Db        304 GCGTTTGCCCCATGCGTAGTCCAGAAGACCCGGACCGCATGATGAAATATGCCAGCAAAC 363

Qy        323 TTAAGAAAA------GTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACCTGC 376
              |   | |||      || | |||||  |||||||| |        |||      ||||| |
Db        364 TGGCGGAAAAAGCATGTAAGATTACAAACAAGAACTTGCATGAGAAGATCAAGGACCTCC 423

Qy        377 TGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACGCCA 436
              |   ||  |    |||| || || || || ||||| || ||  | ||||  || |  |
Db        424 GGACCGTACTTGATACACCGGATGCTGAAACGCCATCACTCTGCTTCCACAACGATGTTA 483

Qy        437 CGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCACCGA 496
              | || |      ||         | ||    || || ||||||   || | || ||
Db        484 CCTGCAACACGCGTGCCGAGTACTCCGTCATGCAGGACGTGTACA---TCAACGCTCCCG 540

Qy        497 CATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTGACA 556
              | | || |||||||||||  | ||||| ||  ||||  | ||||||| ||||| ||||
Db        541 GAACTATTTACCACCAGGCTATGAAAGGCGTGCGGACCCTGTACTGGATTGGCTTCGACA 600

Qy        557 CGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACTGGG 616
              | ||||  || |||| |      ||||||||||| |||||| | |||||||| |||||||
Db        601 CCACCCAGTTCATGTTCTCGGCTATGGCAGGTTCGTACCCTGCATACAACACCAACTGGG 660

Qy        617 CTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGGAGA 676
              | |||||  | |||  | |||||  |||||||| || ||| | | | || | ||    ||
Db        661 CCGACGAAAAAGTCCTTGAAGCGCGTAACATCGGACTCTGCAGCACAAAGCTGAGTGAAG 720

Qy        677 GCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGATCA 736
```

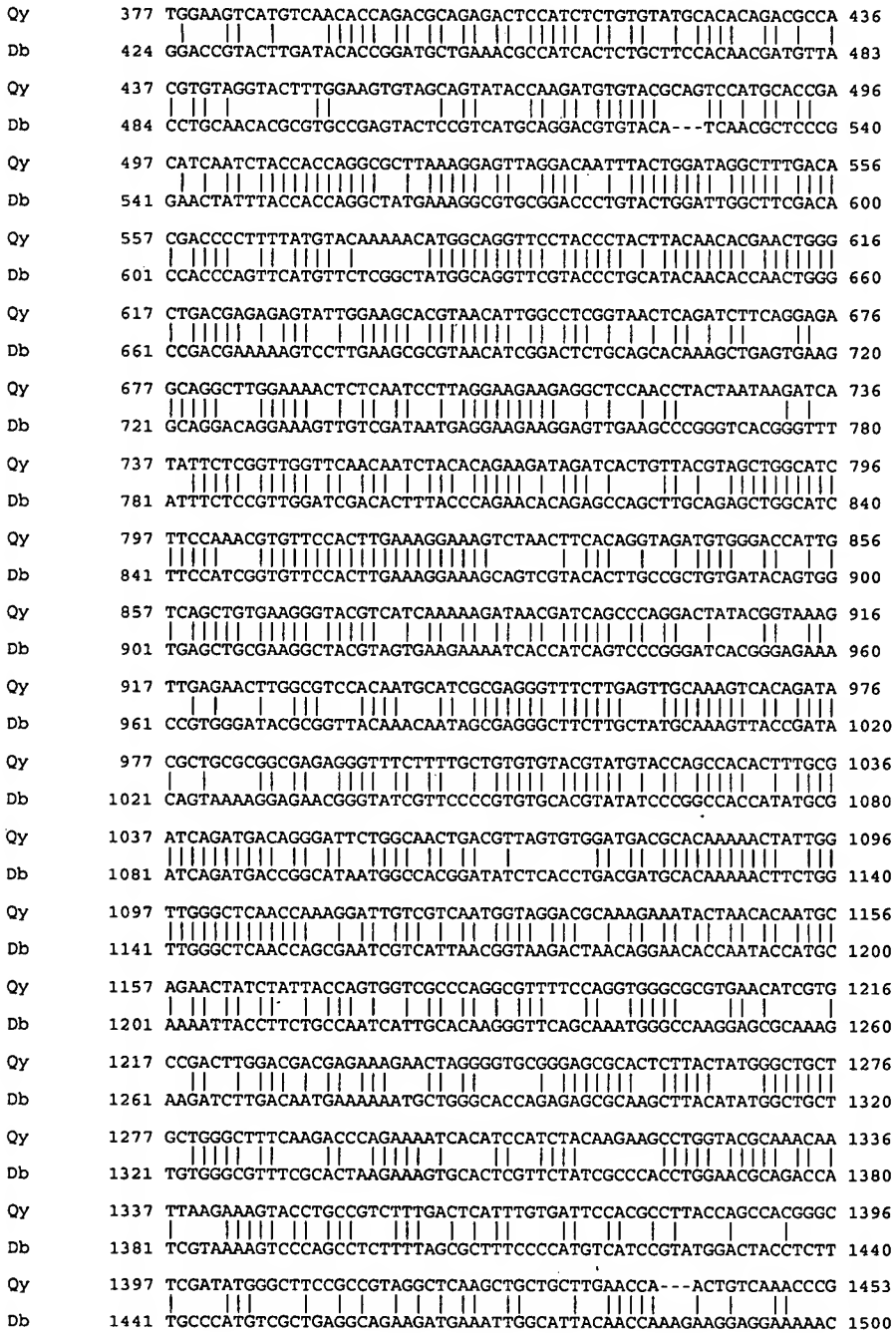
```
                |||||  |||||  | || ||  | |||||||||  | |  | ||         | |
Db       721 GCAGGACAGGAAAGTTGTCGATAATGAGGAAGAAGGAGTTGAAGCCCGGGTCACGGGTTT 780

Qy       737 TATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGCATC 796
               |||||  |||||  || ||| | |||  |||||  | |||  |     || |  ||||||||||
Db       781 ATTTCTCCGTTGGATCGACACTTTACCCAGAACACAGAGCCAGCTTGCAGAGCTGGCATC 840

Qy       797 TTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTG 856
              |||||   |||||||||||||||||||||        | |||  |  | ||||  ||  | |
Db       841 TTCCATCGGTGTTCCACTTGAAAGGAAAGCAGTCGTACACTTGCCGCTGTGATACAGTGG 900

Qy       857 TCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAG 916
              | ||||| ||||| |||||  | || || || || |||||  || ||  |    ||  ||
Db       901 TGAGCTGCGAAGGCTACGTAGTGAAGAAAATCACCATCAGTCCCGGGATCACGGGAGAAA 960

Qy       917 TTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATA 976
              | |   |  |||  ||||  || ||||||| ||||||  |||||||| || ||||
Db       961 CCGTGGGATACGCGGTTACAAACAATAGCGAGGGCTTCTTGCTATGCAAAGTTACCGATA 1020

Qy       977 CGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCG 1036
              | |    || ||  |||| || ||  | ||||| ||||||  | || |||||  | ||||
Db      1021 CAGTAAAAGGAGAACGGGTATCGTTCCCCGTGTGCACGTATATCCCGGCCACCATATGCG 1080

Qy      1037 ATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTATTGG 1096
              |||||||||| || ||  |||| || ||  |    || || ||||||||||  |||
Db      1081 ATCAGATGACCGGCATAATGGCCACGGATATCTCACCTGACGATGCACAAAAACTTCTGG 1140

Qy      1097 TTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAATGC 1156
              |||||||||||||  | || ||| | || |||| |||  | || || || || ||||
Db      1141 TTGGGCTCAACCAGCGAATCGTCATTAACGGTAAGACTAACAGGAACACCAATACCATGC 1200

Qy      1157 AGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATCGTG 1216
              | || || ||  | ||| |  | || || | |||   ||  |||||     || |     |
Db      1201 AAAATTACCTTCTGCCAATCATTGCACAAGGGTTCAGCAAATGGGCCAAGGAGCGCAAAG 1260

Qy      1217 CCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCTGCT 1276
               ||  | ||| | || |||   || ||     | |||||||  |||||     |||||||
Db      1261 AAGATCTTGACAATGAAAAAATGCTGGGCACCAGAGAGCGCAAGCTTACATATGGCTGCT 1320

Qy      1277 GCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCTGGTACGCAAACAA 1336
              |||||  ||    ||  |||||  |    ||  ||||       |||||  |||||  || |
Db      1321 TGTGGGCGTTTCGCACTAAGAAAGTGCACTCGTTCTATCGCCCACCTGGAACGCAGACCA 1380

Qy      1337 TTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCACGGGC 1396
              |    |||||  || |||   |||  | | ||   ||  ||  | |    |  |
Db      1381 TCGTAAAAGTCCCAGCCTCTTTTAGCGCTTTCCCCATGTCATCCGTATGGACTACCTCTT 1440

Qy      1397 TCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCA---ACTGTCAAACCCG 1453
              |   |||    |  | |  | | || ||    |  |||||   |  |  ||
Db      1441 TGCCCATGTCGCTGAGGCAGAAGATGAAATTGGCATTACAACCAAAGAAGGAGGAAAAAC 1500

Qy      1454 CACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAG 1513
              | |    |  |    ||    ||  | |           |  ||| ||| | |
Db      1501 TGCTGCAAGTCCCGGAGGAATTAGTTATGGAGGCCAAGGCTGCTTTCGAGGATGCTCAGG 1560

Qy      1514 AAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAA 1573
              | |   |  ||||||  || || ||||||| || ||||| ||  |  | || | ||  |
Db      1561 AGGAATCCAGAGCGGAGAAGCTCCGAGAAGCACTCCCACCATTAGTGGCAGACAAAGGTA 1620

Qy      1574 AAGAGACCGTAGAGGCAG---------AAGTAGACCTCATTATGCAAGAGGCAGGAGCAG 1624
              |||  |  |  || ||         |||| ||    |   |   ||  | |||||||
Db      1621 TCGAGGCAGCTGCGGAAGTTGTCTGCGAAGTGGAGGGGCTCCAGGCGGACACCGGAGCAG 1680

Qy      1625 GTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAGA 1684
              |||  || || || || || ||  | ||| | |  |  ||    ||     | ||
Db      1681 CACTCGTCGAAACCCCGCGCGGTCATGTAAGGATAATACCTCAAGCAAATGACCGTATGA 1740

Qy      1685 TTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGTA 1744
              | ||    ||     |  | || ||    | ||  |||| |  |  ||||| ||
Db      1741 TCGGACAGTATATCGTTGTCTCGCCGATCTCTGTGCTGAAGAACGCTAAACTCGCACCAG 1800

Qy      1745 TCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATACA 1804
              |||||  | || || || ||   | | || || |||   || ||  |||| || ||
```

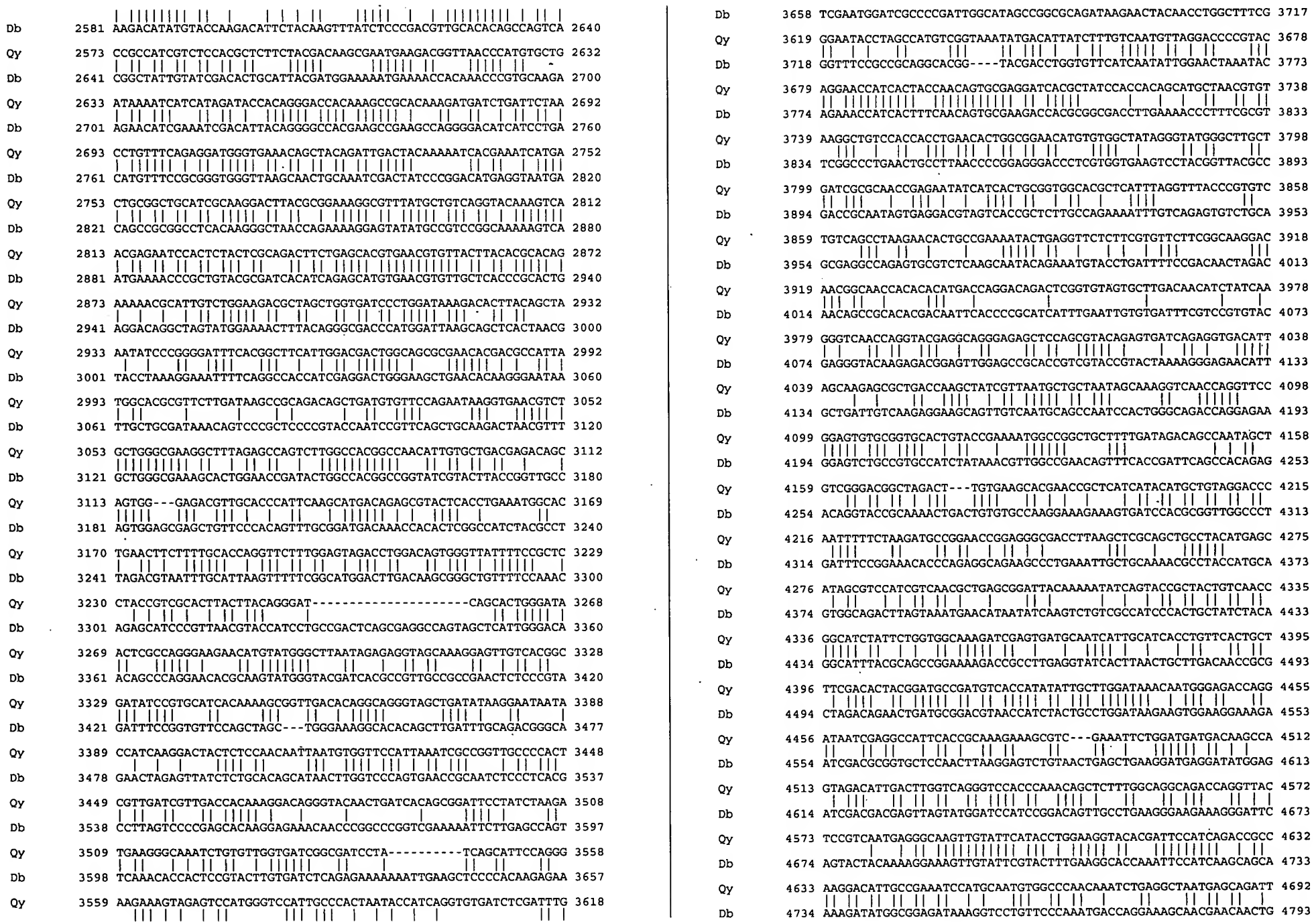

Db 1801 CACACCCGCTAGCAGACCAGGTTAAGATCATAACGCACTCCGGAAGATCAGGAAGGTATG 1860
Qy 1805 AAGTCGAGCCATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAAG 1864
Db 1861 CAGTCGAACCATACGACGCTAAAGTACTGATGCCAGCAGGAAGTGCCGTACCATGGCCAG 1920
Qy 1865 ACTTCCAGGCATTGAGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAACA 1924
Db 1921 AATTCTTAGCACTGAGTGAGAGCGCCACGCTTGTGTACAACGAAAGAGAGTTTGTGAACC 1980
Qy 1925 GATACCTGCACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTATA 1984
Db 1981 GCAAGCTGTACCATATTGCCATGCACGGTCCCGCTAAGAATACAGAAGAGGAGCAGTACA 2040
Qy 1985 AGACTGTAAAGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAGT 2044
Db 2041 AGGTTACAAAGGCAGAGCTCGCAGAAACAGAGTACGTGTTTGACGTGGACAAGAAGCGAT 2100
Qy 2045 GTGTTAAGCGAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCAT 2104
Db 2101 GCGTTAAGAAGGAAGAAGCCTCAGGACTTGTCCTTTCGGGAGAACTGACCAACCCGCCCT 2160
Qy 2105 TTCACGAGTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCAA 2164
Db 2161 ATCACGAACTAGCTCTTGAGGGACTGAAGACTCGACCCGCGGTCCCGTACAAGGTTGAAA 2220
Qy 2165 CCATCGGAGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTGA 2224
Db 2221 CAATAGGAGTGATAGGCACACCAGGATCGGGCAAGTCAGCTATCATCAAGTCAACTGTCA 2280
Qy 2225 CTAAGAAAGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGATG 2284
Db 2281 CGGCACGTGATCTTGTTACCAGCGGAAAGAAAGAAAACTGCCGCGAAATTGAGGCCGACG 2340
Qy 2285 TAAGGAGGATGAGACGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAATG 2344
Db 2341 TGCTACGGCTGAGGGGCATGCAGATCACGTCGAAGACAGTGGATTCGGTTATGCTCAACG 2400
Qy 2345 GGGTTAAGCACCCCGTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGGA 2404
Db 2401 GATGCCACAAAGCCGTAGAAGTGCTGTATGTTGACGAAGCGTTCCGGTGCCACGCAGGAG 2460
Qy 2405 CGCTGCTGGCACTGATTGCCATCGTCAAACC---TAAGAAAGTGGTATTGTGCGGGGACC 2461
Db 2461 CACTACTTGCCTTGATTGCAATCGTCAGACCCCGTAAGAAGGTAGTACTATGCGGAGACC 2520
Qy 2462 CAAAACAATGCGGCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAACC--------- 2512
Db 2521 CTAAGCAATGCGGATTCTTCAACATGATGCAACTAAAGGTACATTTCAACCACCCTGAAA 2580
Qy 2513 ATGACATATGCACTGAAGTGTACCATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAA 2572
Db 2581 AAGACATATGTACCAAGACATTCTACAAGTTTATCTCCCGACGTTGCACACAGCCAGTCA 2640
Qy 2573 CCGCCATCGTCTCCACGCTCTTCTACGACAAGCGAATGAAGACGGTTAACCCATGTGCTG 2632
Db 2641 CGGCTATTGTATCGACACTGCATTACGATGGAAAAATGAAAACCACAAACCCGTGCAAGA 2700
Qy 2633 ATAAAATCATCATAGATACCACAGGGACCACAAAGCCGCACAAAGATGATCTGATTCTAA 2692
Db 2701 AGAACATCGAAATCGACATTACAGGGGCCACGAAGCCGAAGCCAGGGGACATCATCCTGA 2760
Qy 2693 CCTGTTTCAGAGGATGGGTGAAACAGCTACAGATTGACTACAAAAATCACGAAATCATGA 2752
Db 2761 CATGTTTCCGCGGGTGGGTTAAGCAACTGCAAATCGACTATCCCGGACATGAGGTAATGA 2820
Qy 2753 CTGCGGCTGCATCGCAAGGACTTACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCA 2812
Db 2821 CAGCCGCGGCCTCACAAGGGCTAACCAGAAAAGGAGTATATGCCGTCCGGCAAAAAGTCA 2880
Qy 2813 ACGAGAATCCACTCTACTCGCAGACTTCTGAGCACGTGAACGTGTTACTTACACGCACAG 2872
Db 2881 ATGAAAACCCGCTGTACGCGATCACATCAGAGCATGTGAACGTGTTGCTCACCCGCACTG 2940
Qy 2873 AAAAACGCATTGTCTGGAAGACGCTAGCTGGTGATCCCTGGATAAAGACACTTACAGCTA 2932
Db 2941 AGGACAGGCTAGTATGGAAAACTTTACAGGGCGACCCATGGATTAAGCAGCTCACTAACG 3000
Qy 2933 AATATCCCGGGGATTTCACGGCTTCATTGGACGACTGGCAGCGCGAACACGACGCCATTA 2992
Db 3001 TACCTAAAGGAAATTTTCAGGCCACCATCGAGGACTGGGAAGCTGAACACAAGGGAATAA 3060
Qy 2993 TGGCACGCGTTCTTGATAAGCCGCAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCT 3052
Db 3061 TTGCTGCGATAAACAGTCCCGCTCCCCGTACCAATCCGTTCAGCTGCAAGACTAACGTTT 3120
Qy 3053 GCTGGGCGAAGGCTTTAGAGCCAGTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGC 3112
Db 3121 GCTGGGCGAAAGCACTGGAACCGATACTGGCCACGGCCGGTATCGTACTTACCGGTTGCC 3180
Qy 3113 AGTGG---GAGACGTTGCACCCATTCAAGCATGACAGAGCGTACTCACCTGAAATGGCAC 3169
Db 3181 AGTGGAGCGAGCTGTTCCCACAGTTTGCGGATGACAAACCACACTCGGCCATCTACGCCT 3240
Qy 3170 TGAACTTCTTTTGCACCAGGTTCTTTGGAGTAGACCTGGACAGTGGGTTATTTTCCGCTC 3229
Db 3241 TAGACGTAATTTGCATTAAGTTTTTCGGCATGGACTTGACAAGCGGGCTGTTTTCCAAAC 3300
Qy 3230 CTACCGTCGCACTTACTTACAGGGAT----------CAGCACTGGGATA 3268
Db 3301 AGAGCATCCCGTTAACGTACCATCCTGCCGACTCAGCGAGGCCAGTAGCTCATTGGGACA 3360
Qy 3269 ACTCGCCAGGGAAGAACATGTATGGGCTTAATAGAGAGGTAGCAAAGGAGTTGTCACGGC 3328
Db 3361 ACAGCCCAGGAACACGCAAGTATGGGTACGATCACGCCGTTGCCGCCGAACTCTCCCGTA 3420
Qy 3329 GATATCCGTGCATCACAAAAGCGGTTGACACAGGCAGGGTAGCTGATATAAGGAATAATA 3388
Db 3421 GATTTCCGGTGTTCCAGCTAGC---TGGGAAAGGCACACAGCTTGATTTGCAGACGGGCA 3477
Qy 3389 CCATCAAGGACTACTCTCCAACAATTAATGTGGTTCCATTAAATCGCCGGTTGCCCCACT 3448
Db 3478 GAACTAGAGTTATCTCTGCACAGCATAACTTGGTCCCAGTGAACCGCAATCTCCCTCACG 3537
Qy 3449 CGTTGATCGTTGACCACAAAGGACAGGGTACAACTGATCACAGCGGATTCCTATCTAAGA 3508
Db 3538 CCTTAGTCCCCGAGCACAAGGAGAAACAACCCGGCCCGGTCGAAAAATTCTTGAGCCAGT 3597
Qy 3509 TGAAGGGCAAATCTGTGTTGGTGATCGGCGATCCTA----------TCAGCATTCCAGGG 3558
Db 3598 TCAAACACCACTCCGTACTTGTGATCTCAGAGAAAAAAATTGAAGCTCCCCACAAGAGAA 3657
Qy 3559 AAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATACCATCAGGTGTGATCTCGATTTG 3618
Db 3658 TCGAATGGATCGCCCCGATTGGCATAGCCGGCGCAGATAAGAACTACAACCTGGCTTTCG 3717
Qy 3619 GGAATACCTAGCCATGTCGGTAAATATGACATTATCTTTGTCAATGTTAGGACCCCGTAC 3678
Db 3718 GGTTTCCGCCGCAGGCACGG----TACGACCTGGTGTTCATCAATATTGGAACTAAATAC 3773
Qy 3679 AGGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCACCACAGCATGCTAACGTGT 3738
Db 3774 AGAAACCATCACTTTCAACAGTGCGAAGACCACGCGGCGACCTTGAAAACCCTTTCGCGT 3833
Qy 3739 AAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCTATAGGGTATGGGCTTGCT 3798
Db 3834 TCGGCCCTGAACTGCCTTAACCCCGGAGGCACCCTCGTGGTGAAGTCCTACGGTTACGCC 3893
Qy 3799 GATCGCGCAACCGAGAATATCATCACTGCGGTGGCACGCTCATTTAGGTTTACCCGTGTC 3858
Db 3894 GACCGCAATAGTGAGGACGTAGTCACCGCTCTTGCCAGAAAATTTGTCAGAGTGTCTGCA 3953
Qy 3859 TGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTCGTGTTCTTCGGCAAGGAC 3918
Db 3954 GCGAGGCCAGAGTGCGTCTCAAGCAATACAGAAATGTACCTGATTTTCCGACAACTAGAC 4013



```
RESULT 12
US-09-102-248-7
; Sequence 7, Application US/09102248
; Patent No. 6008035
;  GENERAL INFORMATION:
;    APPLICANT:  Johnston, Robert E.
;    APPLICANT:  Davis, Nancy L.
;    APPLICANT:  Simpson, Dennis A.
;    TITLE OF INVENTION:  System for the In Vivo Delivery and
;    TITLE OF INVENTION:  Expression of Heterologous Genes in the Bone Marrow
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson, P.A.
;      STREET:  1211 East Morehead Street
;      CITY:  Charlotte
;      STATE:  No. 6008035th Carolina
;      COUNTRY:  USA
;      ZIP:  28234
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/102,248
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/801,263
;      FILING DATE:  19-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sibley, Kenneth D.
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5470-147
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-420-2200
;      TELEFAX:  919-881-3175
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11663 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-102-248-7

  Query Match             26.7%;  Score 1333.2;  DB 3;  Length 11663;
  Best Local Similarity   56.1%;  Pred. No. 0;
  Matches 2820;  Conservative    0;  Mismatches 2128;  Indels   80;  Gaps   13;
```



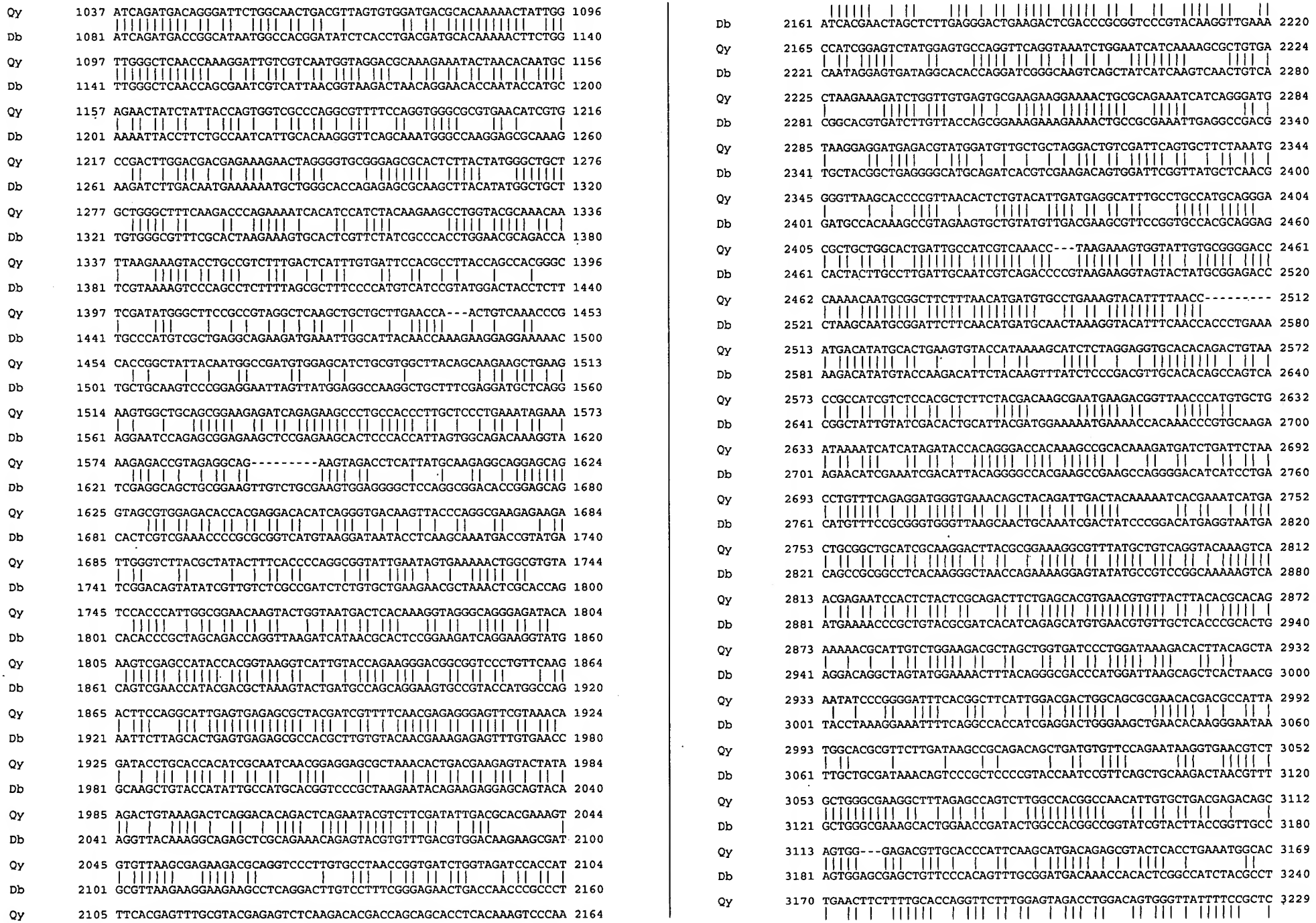

Qy 377 TGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACGCCA 436
Db 424 GGACCGTACTTGATACACCGGATGCTGAAACGCCATCACTCTGCTTCCACAACGATGTTA 483
Qy 437 CGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCACCGA 496
Db 484 CCTGCAACACGCGTGCCGAGTACTCCGTCATGCAGGACGTGTACA---TCAACGCTCCCG 540
Qy 497 CATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTGACA 556
Db 541 GAACTATTTACCACCAGGCTATGAAAGGCGTGCGGACCCTGTACTGGATTGGCTTCGACA 600
Qy 557 CGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACTGGG 616
Db 601 CCACCCAGTTCATGTTCTCGGCTATGGCAGGTTCGTACCCTGCATACAACACCAACTGGG 660
Qy 617 CTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGGAGA 676
Db 661 CCGACGAAAAAGTCCTTGAAGCGCGTAACATCGGACTCTGCAGCACAAAGCTGAGTGAAG 720
Qy 677 GCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGATCA 736
Db 721 GCAGGACAGGAAAGTTGTCGATAATGAGGAAGAAGGAGTTGAAGCCCGGGTCACGGGTTT 780
Qy 737 TATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGCATC 796
Db 781 ATTTCTCCGTTGGATCGACACTTTACCCAGAACACAGAGCCAGCTTGCAGAGCTGGCATC 840
Qy 797 TTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTG 856
Db 841 TTCCATCGGTGTTCCACTTGAAAGGAAAGCAGTCGTACACTTGCCGCTGTGATACAGTGG 900
Qy 857 TCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAG 916
Db 901 TGAGCTGCGAAGGCTACGTAGTGAAGAAAATCACCATCAGTCCCGGGATCACGGGAGAAA 960
Qy 917 TTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATA 976
Db 961 CCGTGGGATACGCGGTTACAAACAATAGCGAGGGCTTCTTGCTATGCAAAGTTACCGATA 1020
Qy 977 CGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCG 1036
Db 1021 CAGTAAAAGGAGAACGGGTATCGTTCCCCGTGTGCACGTATATCCCGGCCACCATATGCG 1080
Qy 1037 ATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTATTGG 1096
Db 1081 ATCAGATGACCGGCATAATGGCCACGGATATCTCACCTGACGATGCACAAAAACTTCTGG 1140
Qy 1097 TTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAATGC 1156
Db 1141 TTGGGCTCAACCAGCGAATCGTCATTAACGGTAAGACTAACAGGAACACCAATACCATGC 1200
Qy 1157 AGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATCGTG 1216
Db 1201 AAAATTACCTTCTGCCAATCATTGCACAAGGGTTCAGCAAATGGGCCAAGGAGCGCAAAG 1260
Qy 1217 CCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCTGCT 1276
Db 1261 AAGATCTTGACAATGAAAAAATGCTGGGCACCAGAGAGCGCAAGCTTACATATGGCTGCT 1320
Qy 1277 GCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCTGGTACGCAAACAA 1336
Db 1321 TGTGGGCGTTTCGCACTAAGAAAGTGCACTCGTTCTATCGCCCACCTGGAACGCAGACCA 1380
Qy 1337 TTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCACGGGC 1396
Db 1381 TCGTAAAAGTCCCAGCCTCTTTTAGCGCTTTCCCCATGTCATCCGTATGGACTACCTCTT 1440
Qy 1397 TCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCA---ACTGTCAAACCCG 1453
Db 1441 TGCCCATGTCGCTGAGGCAGAAGATGAAATTGGCATTACAACCAAAGAAGGAGGAAAAAC 1500
Qy 1454 CACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAG 1513
Db 1501 TGCTGCAAGTCCCGGAGGAATTAGTTATGGAGGCCAAGGCTGCTTTCGAGGATGCTCAGG 1560
Qy 1514 AAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAA 1573
Db 1561 AGGAATCCAGAGCGGAGAAGCTCCGAGAAGCACTCCCACCATTAGTGGCAGACAAAGGTA 1620
Qy 1574 AAGAGACCGTAGAGGCAG---------AAGTAGACCTCATTATGCAAGAGGCAGGAGCAG 1624
Db 1621 TCGAGGCAGCTGCGGAAGTTGTCTGCGAAGTGGAGGGGCTCCAGGCGGACACCGGAGCAG 1680
Qy 1625 GTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAGA 1684
Db 1681 CACTCGTCGAAACCCCGCGCGGTCATGTAAGGATAATACCTCAAGCAAATGACCGTATGA 1740
Qy 1685 TTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGTA 1744
Db 1741 TCGGACAGTATATCGTTGTCTCGCCGATCTCTGTGCTGAAGAACGCTAAACTCGCACCAG 1800
Qy 1745 TCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATACA 1804
Db 1801 CACACCCGCTAGCAGACCAGGTTAAGATCATAACGCACTCCGGAAGATCAGGAAGGTATG 1860
Qy 1805 AAGTCGAGCCATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAAG 1864
Db 1861 CAGTCGAACCATACGACGCTAAAGTACTGATGCCAGCAGGAAGTGCCGTACCATGGCCAG 1920
Qy 1865 ACTTCCAGGCATTGAGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAACA 1924
Db 1921 AATTCTTAGCACTGAGTGAGAGCGCCACGCTTGTGTACAACGAAAGAGAGTTTGTGAACC 1980
Qy 1925 GATACCTGCACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTATA 1984
Db 1981 GCAAGCTGTACCATATTGCCATGCACGGTCCCGCTAAGAATACAGAAGAGGAGCAGTACA 2040
Qy 1985 AGACTGTAAAGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAGT 2044
Db 2041 AGGTTACAAAGGCAGAGCTCGCAGAAACAGAGTACGTGTTTGACGTGGACAAGAAGCGAT 2100
Qy 2045 GTGTTAAGCGAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCAT 2104
Db 2101 GCGTTAAGAAGGAAGAAGCCTCAGGACTTGTCCTTTCGGGAGAACTGACCAACCCGCCCT 2160
Qy 2105 TTCACGAGTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCAA 2164
Db 2161 ATCACGAACTAGCTCTTGAGGGACTGAAGACTCGACCCGCGGTCCCGTACAAGGTTGAAA 2220
Qy 2165 CCATCGGAGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTGA 2224
Db 2221 CAATAGGAGTGATAGGCACACCAGGATCGGGCAAGTCAGCTATCATCAAGTCAACTGTCA 2280
Qy 2225 CTAAGAAAGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGATG 2284
Db 2281 CGGCACGTGATCTTGTTACCAGCGGAAAGAAAGAAAACTGCCGCGAAATTGAGGCCGACG 2340
Qy 2285 TAAGGAGGATGAGACGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAATG 2344
Db 2341 TGCTACGGCTGAGGGGCATGCAGATCACGTCGAAGACAGTGGATTCGGTTATGCTCAACG 2400
Qy 2345 GGGTTAAGCACCCCGTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGGA 2404
Db 2401 GATGCCACAAAGCCGTAGAAGTGCTGTATGTTGACGAAGCGTTCCGGTGCCACGCAGGAG 2460
Qy 2405 CGCTGCTGGCACTGATTGCCATCGTCAAACC---TAAGAAAGTGGTATTGTGCGGGGACC 2461
Db 2461 CACTACTTGCCTTGATTGCAATCGTCAGACCCCGTAAGAAGGTAGTACTATGCGGAGACC 2520
Qy 2462 CAAAACAATGCGGCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAACC--------- 2512
Db 2521 CTAAGCAATGCGGATTCTTCAACATGATGCAACTAAAGGTACATTTCAACCACCCTGAAA 2580
Qy 2513 ATGACATATGCACTGAAGTGTACCATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAA 2572

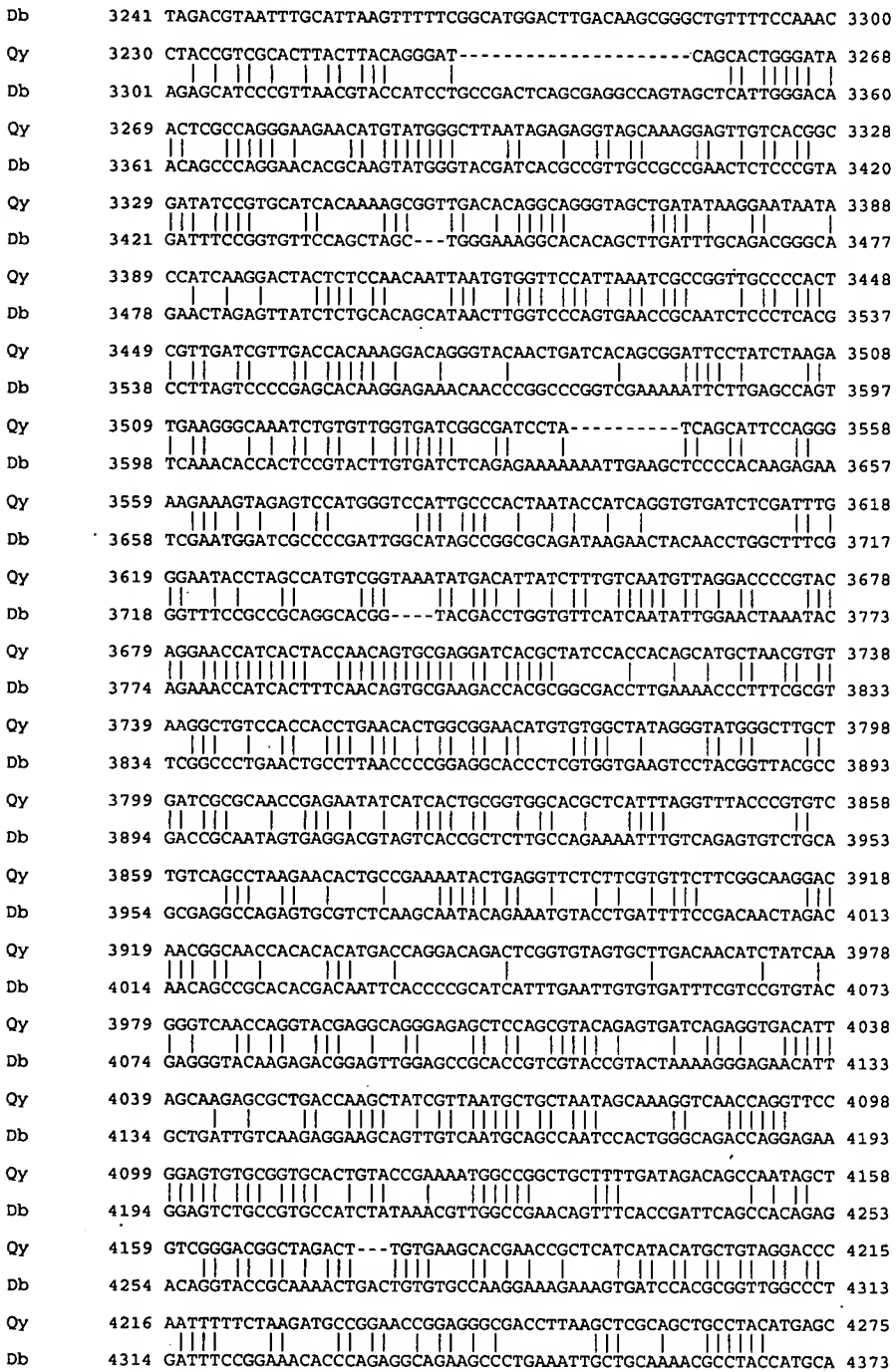


```
Qy       4693 TGCTTGTACATCCTGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGCCCAGTAGAGGAG 4752
Db       4794 TGTGCCTACATATTGGGGGAGACCATGGAAGCAATCCGCGAAAAATGCCCGGTCGACCAC 4853

Qy       4753 TCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTACGCTATGACGGCT 4812
Db       4854 AACCCGTCGTCTAGCCCGCCAAAAACGCTGCCGTGCCTCTGTATGTATGCCATGACGCCA 4913

Qy       4813 GAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTATGCTCATCATTC 4872
Db       4914 GAAAGGGTCCACAGACTCAGAAGCAATAACGTCAAAGAAGTTACAGTATGCTCCTCCACC 4973

Qy       4873 CTGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGCAAACCAGTCCTG 4932
Db       4974 CCCCTTCCAAAGTACAAAATCAAGAATGTTCAGAAGGTTCAGTGCACAAAAGTAGTCCTG 5033

Qy       4933 TTTTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGCGGAA 4980
Db       5034 TTTAACCCGCATACCCCCGCATTCGTTCCCGCCCGTAAGTACATAGAA 5081
```

```
RESULT 13
US-09-367-764-1
; Sequence 1, Application US/09367764
; Patent No. 6583121
;  GENERAL INFORMATION:
;    APPLICANT:  Johnston, Robert E.
;    APPLICANT:  Davis, Nancy L.
;    APPLICANT:  Simpson, Dennis A.
;    TITLE OF INVENTION:  System for the In Vivo Delivery and
;    TITLE OF INVENTION:  Expression of Heterologous Genes in the Bone Marrow
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson, P.A.
;      STREET:  1211 East Morehead Street
;      CITY:  Charlotte
;      STATE:  No. 6583121th Carolina
;      COUNTRY:  USA
;      ZIP:  28234
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/367,764
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/801,263
;      FILING DATE:  19-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sibley, Kenneth D.
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5470-147
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-420-2200
;      TELEFAX:  919-881-3175
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11663 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  60..7559
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  7608..11342
US-09-367-764-1
```

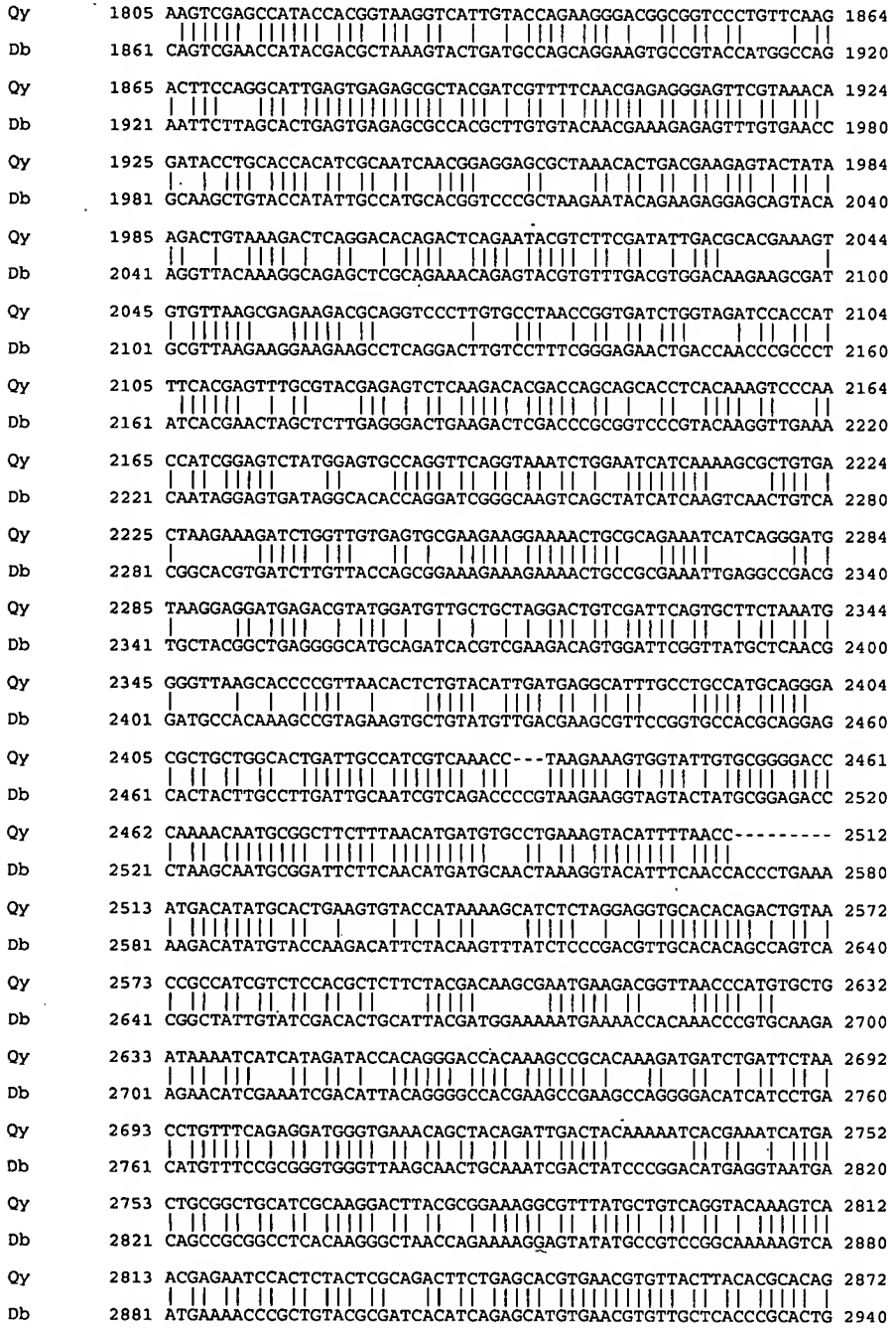
```
  Query Match           26.7%;  Score 1333.2;  DB 4;  Length 11663;
  Best Local Similarity  56.1%;  Pred. No. 0;
  Matches 2820;  Conservative    0;  Mismatches 2128;  Indels   80;  Gaps   13;

Qy         23 ATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGTCGTTACAGC 82
Db         64 AGAAGCCAGTAGTTAACGTAGACGTAGACCCTCAGAGTCCGTTTGTCGTGCAACTGCAAA 123

Qy         83 GGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACCATGCCAATG 142
Db        124 AGAGCTTCCCGCAATTTGAGGTAGTAGCACAGCAGGTCACTCCAAATGACCATGCTAATG 183

Qy        143 CCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACCGGGACCAAG 202
Db        184 CCAGAGCATTTTCGCATCTGGCCAGTAAACTAATCGAGCTGGAGGTTCCTACCACAGCGA 243

Qy        203 TTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACCGCTATCATT 262
Db        244 CGATTTTGGACATAGGCAGCGCACCGGCTCGTAGAATGTTTTCCGAGCACCAGTACCATT 303

Qy        263 GTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGGTATGCAGAAAGAC 322
Db        304 GCGTTTGCCCCATGCGTAGTCCAGAAGACCCGGACCGCATGATGAAATATGCCAGCAAAC 363

Qy        323 TTAAGAAAA------GTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACCTGC 376
Db        364 TGGCGGAAAAAGCATGTAAGATTACAAACAAGAACTTGCATGAGAAGATCAAGGACCTCC 423

Qy        377 TGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACGCCA 436
Db        424 GGACCGTACTTGATACACCGGATGCTGAAACGCCATCACTCTGCTTCCACAACGATGTTA 483

Qy        437 CGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCACCGA 496
Db        484 CCTGCAACACGCGTGCCGAGTACTCCGTCATGCAGGACGTGTACA---TCAACGCTCCCG 540

Qy        497 CATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTGACA 556
Db        541 GAACTATTTACCACCAGGCTATGAAAGGCGTGCGGACCCTGTACTGGATTGGCTTCGACA 600

Qy        557 CGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACTGGG 616
Db        601 CCACCCAGTTCATGTTCTCGGCTATGGCAGGTTCGTACCCTGCATACAACACCAACTGGG 660

Qy        617 CTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGGAGA 676
Db        661 CCGACGAAAAAGTCCTTGAAGCGCGTAACATCGGACTCTGCAGCACAAAGCTGAGTGAAG 720

Qy        677 GCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGATCA 736
Db        721 GCAGGACAGGAAAGTTGTCGATAATGAGGAAGAAGGAGTTGAAGCCCGGGTCACGGGTTT 780

Qy        737 TATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGCATC 796
Db        781 ATTTCTCCGTTGGATCGACACTTTACCCAGAACACAGAGCCAGCTTGCAGAGCTGGCATC 840

Qy        797 TTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTG 856
Db        841 TTCCATCGGTGTTCCACTTGAAAGGAAAGCAGTCGTACACTTGCCGCTGTGATACAGTGG 900

Qy        857 TCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAG 916
Db        901 TGAGCTGCGAAGGCTACGTAGTGAAGAAAATCACCATCAGTCCCGGGATCACGGGAGAAA 960

Qy        917 TTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATA 976
Db        961 CCGTGGGATACGCGGTTACAAACAATAGCGAGGGCTTCTTGCTATGCAAAGTTACCGATA 1020

Qy        977 CGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCG 1036
Db       1021 CAGTAAAAGGAGAACGGGTATCGTTCCCCGTGTGCACGTATATCCCGGCCACCATATGCG 1080
```

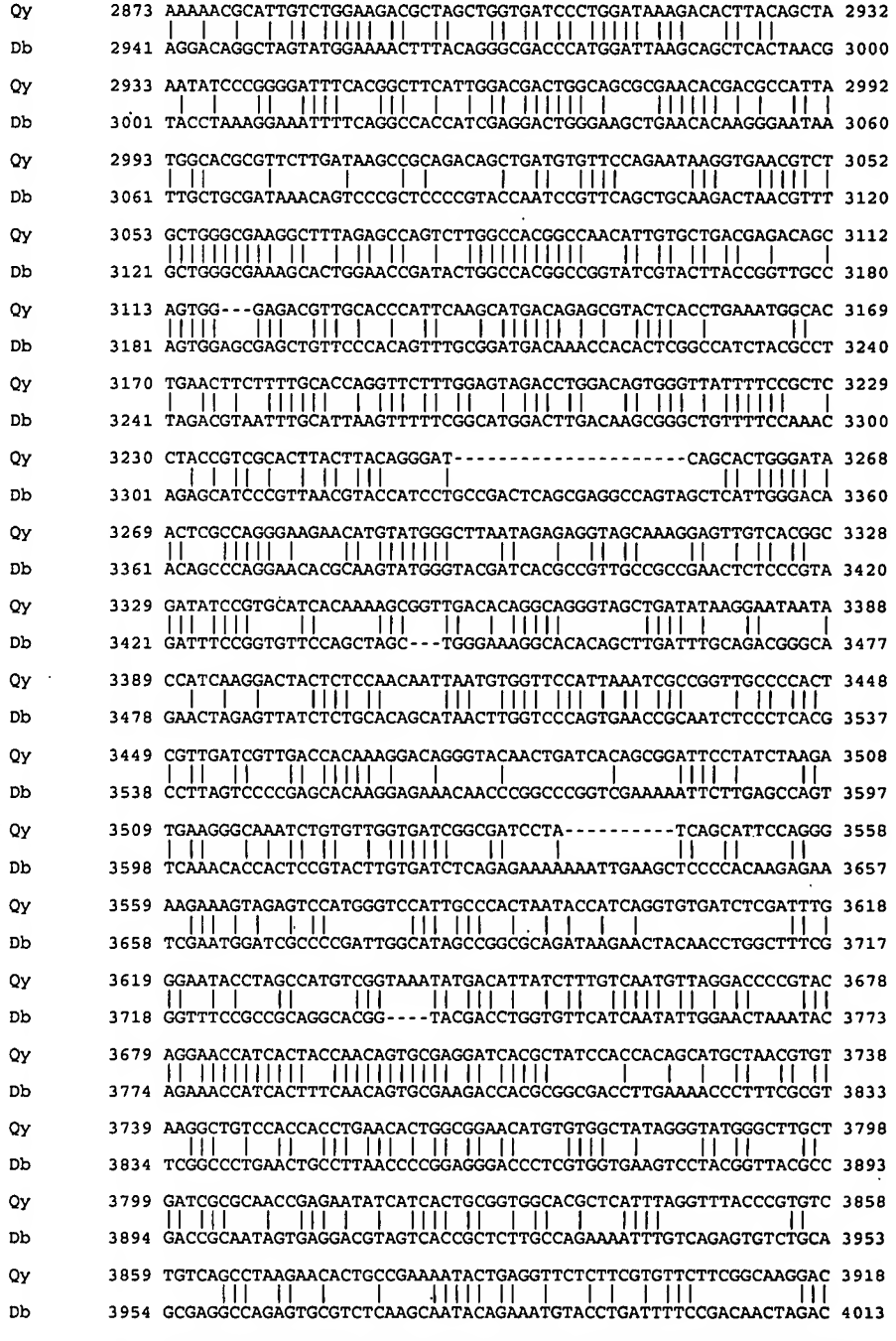


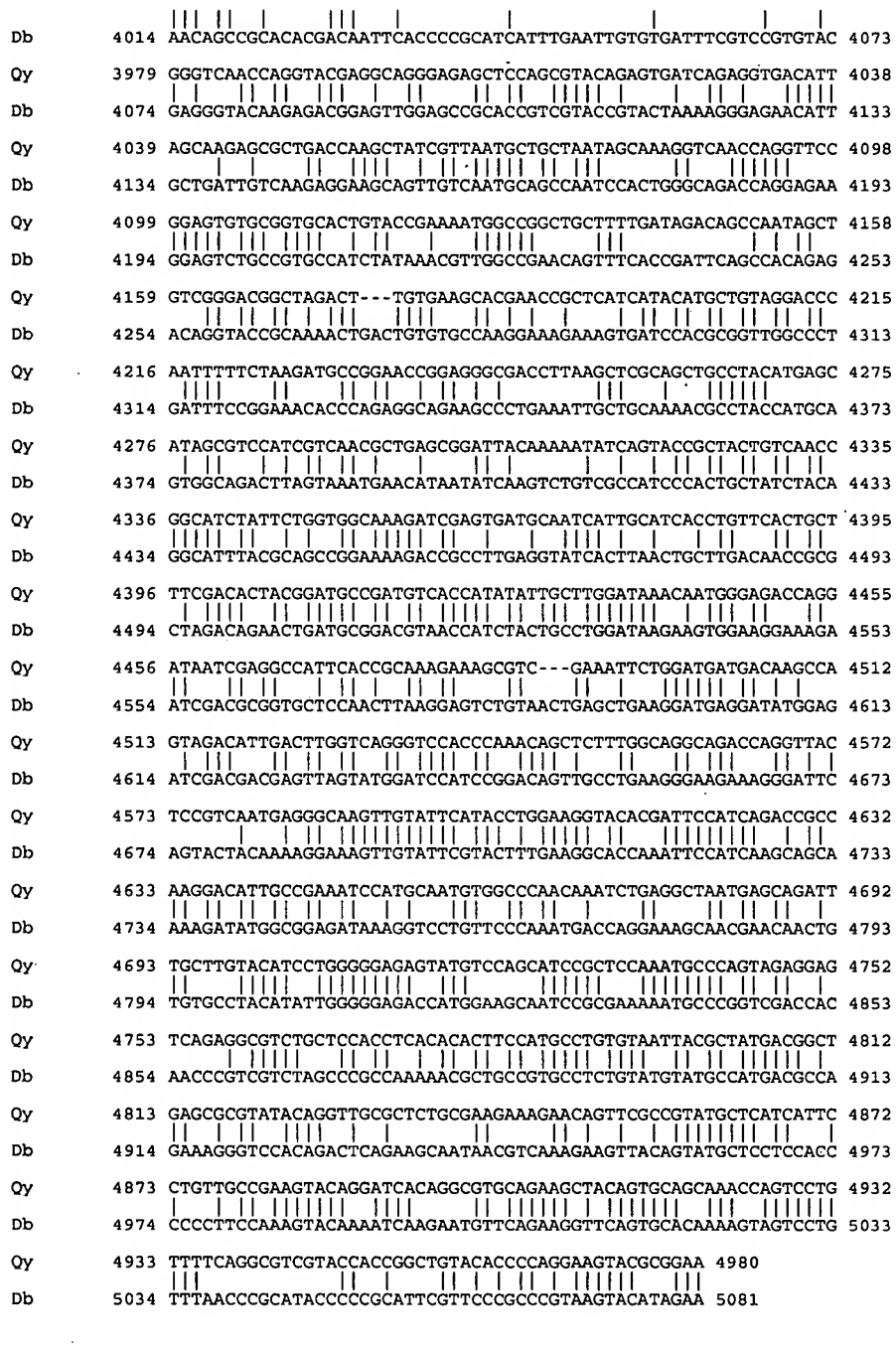



```
RESULT 14
US-09-367-764-7
; Sequence 7, Application US/09367764
; Patent No. 6583121
;  GENERAL INFORMATION:
;    APPLICANT:  Johnston, Robert E.
;    APPLICANT:  Davis, Nancy L.
;    APPLICANT:  Simpson, Dennis A.
;    TITLE OF INVENTION:  System for the In Vivo Delivery and
;    TITLE OF INVENTION:  Expression of Heterologous Genes in the Bone Marrow
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson, P.A.
;      STREET:  1211 East Morehead Street
;      CITY:  Charlotte
;      STATE:  No. 6583121th Carolina
;      COUNTRY:  USA
;      ZIP:  28234
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/367,764
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/801,263
;      FILING DATE:  19-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sibley, Kenneth D.
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5470-147
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-420-2200
;      TELEFAX:  919-881-3175
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11663 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-367-764-7

  Query Match             26.7%;  Score 1333.2;  DB 4;  Length 11663;
  Best Local Similarity   56.1%;  Pred. No. 0;
  Matches 2820;  Conservative    0;  Mismatches 2128;  Indels   80;  Gaps  13;

Qy         23 ATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGTCGTTACAGC 82
Db         64 AGAAGCCAGTAGTTAACGTAGACGTAGACCCTCAGAGTCCGTTTGTCGTGCAACTGCAAA 123

Qy         83 GGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACCATGCCAATG 142
Db        124 AGAGCTTCCCGCAATTTGAGGTAGTAGCACAGCAGGTCACTCCAAATGACCATGCTAATG 183

Qy        143 CCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACCGGGACCAAG 202
Db        184 CCAGAGCATTTTCGCATCTGGCCAGTAAACTGATCGAGCTGGAGGTTCCTACCACAGCGA 243

Qy        203 TTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACCGCTATCATT 262
Db        244 CGATTTTGGACATAGGCAGCGCACCGGCTCGTAGAATGTTTTCCGAGCACCAGTACCATT 303

Qy        263 GTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGGTATGCAGAAAGAC 322
Db        304 GCGTTTGCCCCATGCGTAGTCCAGAAGACCCGGACCGCATGATGAAATATGCCAGCAAAC 363

Qy        323 TTAAGAAAA------GTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACCTGC 376
Db        364 TGGCGGAAAAAGCATGTAAGATTACAAACAAGAACTTGCATGAGAAGATCAAGGACCTCC 423

Qy        377 TGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACGCCA 436
Db        424 GGACCGTACTTGATACACCGGATGCTGAAACGCCATCACTCTGCTTCCACAACGATGTTA 483

Qy        437 CGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCACCGA 496
Db        484 CCTGCAACACGCGTGCCGAGTACTCCGTCATGCAGGACGTGTACA---TCAACGCTCCCG 540

Qy        497 CATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTGACA 556
Db        541 GAACTATTTACCACCAGGCTATGAAAGGCGTGCGGACCCTGTACTGGATTGGCTTCGACA 600

Qy        557 CGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACTGGG 616
Db        601 CCACCCAGTTCATGTTCTCGGCTATGGCAGGTTCGTACCCTGCATACAACACCAACTGGG 660

Qy        617 CTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGGAGA 676
Db        661 CCGACGAAAAAGTCCTTGAAGCGCGTAACATCGGACTCTGCAGCACAAAGCTGAGTGAAG 720

Qy        677 GCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGATCA 736

Db        721 GCAGGACAGGAAAGTTGTCGATAATGAGGAAGAAGGAGTTGAAGCCCGGGTCACGGGTTT 780

Qy        737 TATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGCATC 796
Db        781 ATTTCTCCGTTGGATCGACACTTTACCCAGAACACAGAGCCAGCTTGCAGAGCTGGCATC 840

Qy        797 TTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTG 856
Db        841 TTCCATCGGTGTTCCACTTGAAAGGAAAGCAGTCGTACACTTGCCGCTGTGATACAGTGG 900

Qy        857 TCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAG 916
Db        901 TGAGCTGCGAAGGCTACGTAGTGAAGAAAATCACCATCAGTCCCGGGATCACGGGAGAAA 960

Qy        917 TTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATA 976
Db        961 CCGTGGGATACGCGGTTACAAACAATAGCGAGGGCTTCTTGCTATGCAAAGTTACCGATA 1020

Qy        977 CGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCG 1036
Db       1021 CAGTAAAAGGAGAACGGGTATCGTTCCCCGTGTGCACGTATATCCCGGCCACCATATGCG 1080

Qy       1037 ATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTATTGG 1096
Db       1081 ATCAGATGACCGGCATAATGGCCACGGATATCTCACCTGACGATGCACAAAAACTTCTGG 1140

Qy       1097 TTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAATGC 1156
Db       1141 TTGGGCTCAACCAGCGAATCGTCATTAACGGTAAGACTAACAGGAACACCAATACCATGC 1200

Qy       1157 AGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATCGTG 1216
Db       1201 AAAATTACCTTCTGCCAATCATTGCACAAGGGTTCAGCAAATGGGCCAAGGAGCGCAAAG 1260

Qy       1217 CCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCTGCT 1276
Db       1261 AAGATCTTGACAATGAAAAAATGCTGGGCACCAGAGAGCGCAAGCTTACATATGGCTGCT 1320

Qy       1277 GCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCTGGTACGCAAACAA 1336
Db       1321 TGTGGGCGTTTCGCACTAAGAAAGTGCACTCGTTCTATCGCCCACCTGGAACGCAGACCA 1380

Qy       1337 TTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCACGGGC 1396
Db       1381 TCGTAAAAGTCCCAGCCTCTTTTAGCGCTTTCCCCATGTCATCCGTATGGACTACCTCTT 1440

Qy       1397 TCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCA---ACTGTCAAACCCG 1453
Db       1441 TGCCCATGTCGCTGAGGCAGAAGATGAAATTGGCATTACAACCAAAGAAGGAGGAAAAAC 1500

Qy       1454 CACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAG 1513
Db       1501 TGCTGCAAGTCCCGGAGGAATTAGTTATGGAGGCCAAGGCTGCTTTCGAGGATGCTCAGG 1560

Qy       1514 AAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAA 1573
Db       1561 AGGAATCCAGAGCGGAGAAGCTCCGAGAAGCACTCCCACCATTAGTGGCAGACAAAGGTA 1620

Qy       1574 AAGAGACCGTAGAGGCAG---------AAGTAGACCTCATTATGCAAGAGGCAGGAGCAG 1624
Db       1621 TCGAGGCAGCTGCGGAAGTTGTCTGCGAAGTGGAGGGGCTCCAGGCGGACACCGGAGCAG 1680

Qy       1625 GTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAGA 1684
Db       1681 CACTCGTCGAAACCCCGCGCGGTCATGTAAGGATAATACCTCAAGCAAATGACCGTATGA 1740

Qy       1685 TTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGTA 1744
Db       1741 TCGGACAGTATATCGTTGTCTCGCCGATCTCTGTGCTGAAGAACGCTAAACTCGCACCAG 1800

Qy       1745 TCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATACA 1804
Db       1801 CACACCCGCTAGCAGACCAGGTTAAGATCATAACGCACTCCGGAAGATCAGGAAGGTATG 1860
```



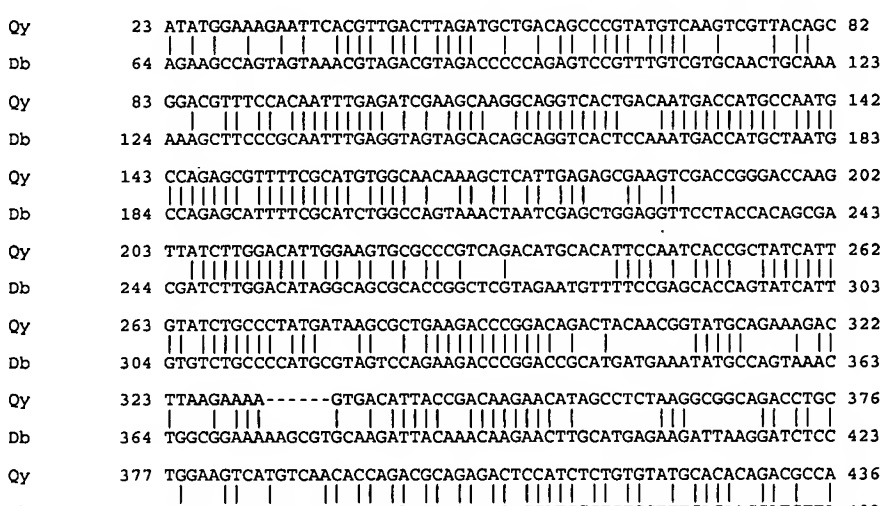


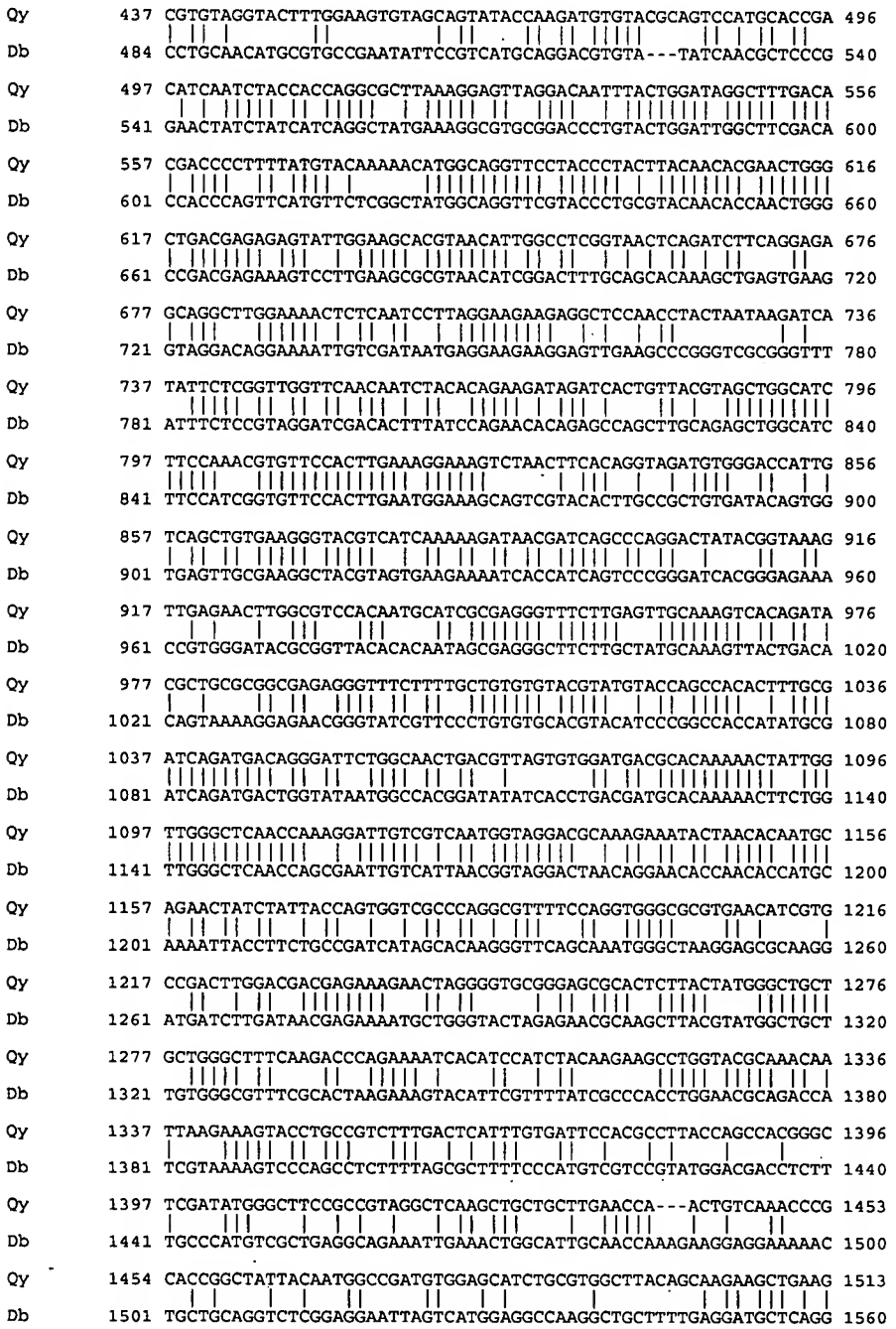


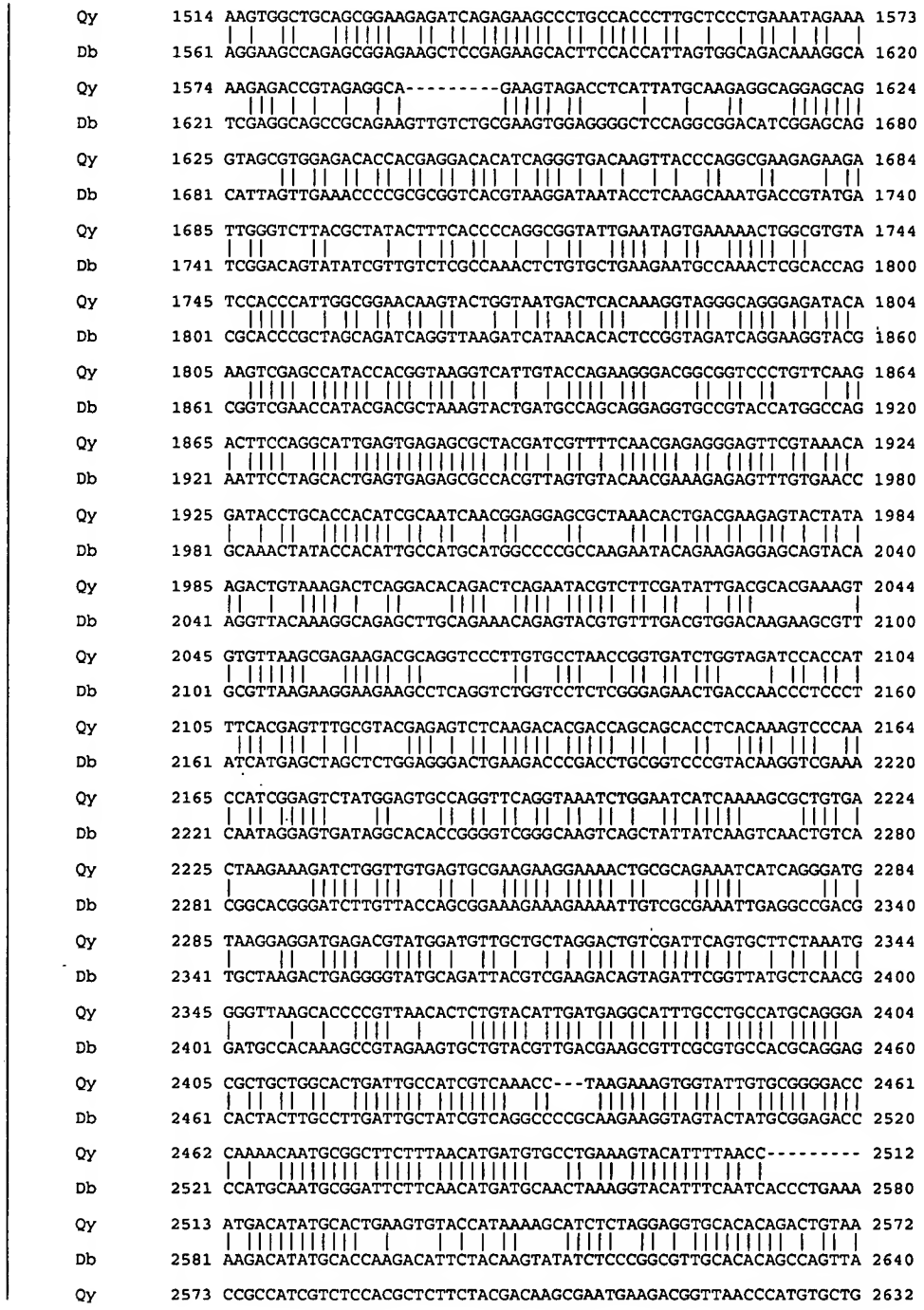
RESULT 15

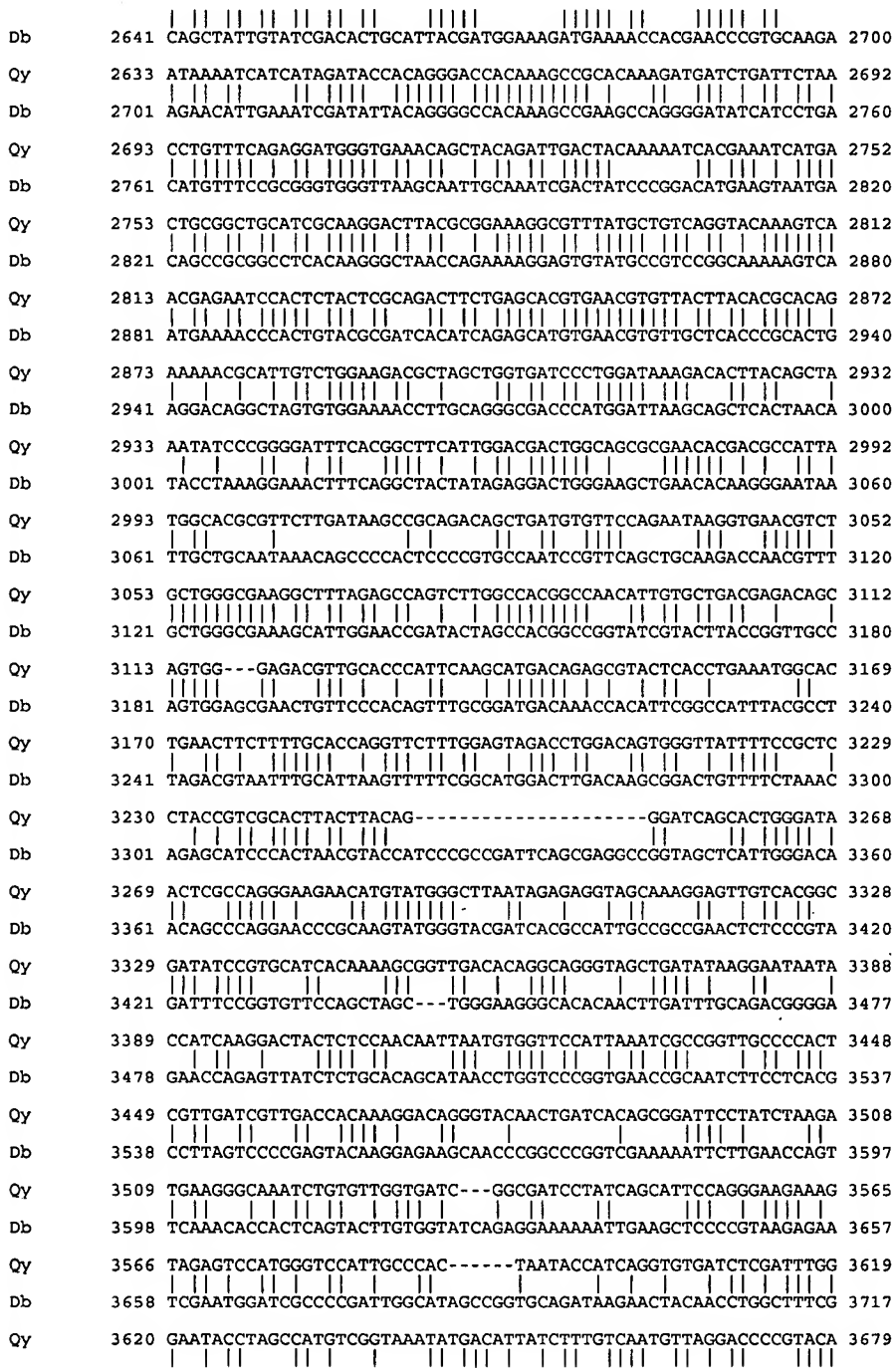
US-08-801-263A-8
; Sequence 8, Application US/08801263A
; Patent No. 5811407
; GENERAL INFORMATION:
; APPLICANT: Johnston, Robert E.
; APPLICANT: Davis, Nancy L.
; APPLICANT: Simpson, Dennis A.
; TITLE OF INVENTION: System for the In Vivo Delivery and
; TITLE OF INVENTION: Expression of Heterologous Genes in the Bone Marrow
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson, P.A.
; STREET: 1211 East Morehead Street
; CITY: Charlotte
; STATE: No. 5811407th Carolina
; COUNTRY: USA
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,263A
; FILING DATE: 19-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5470-147
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-420-2200
; TELEFAX: 919-881-3175
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11703 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-801-263A-8

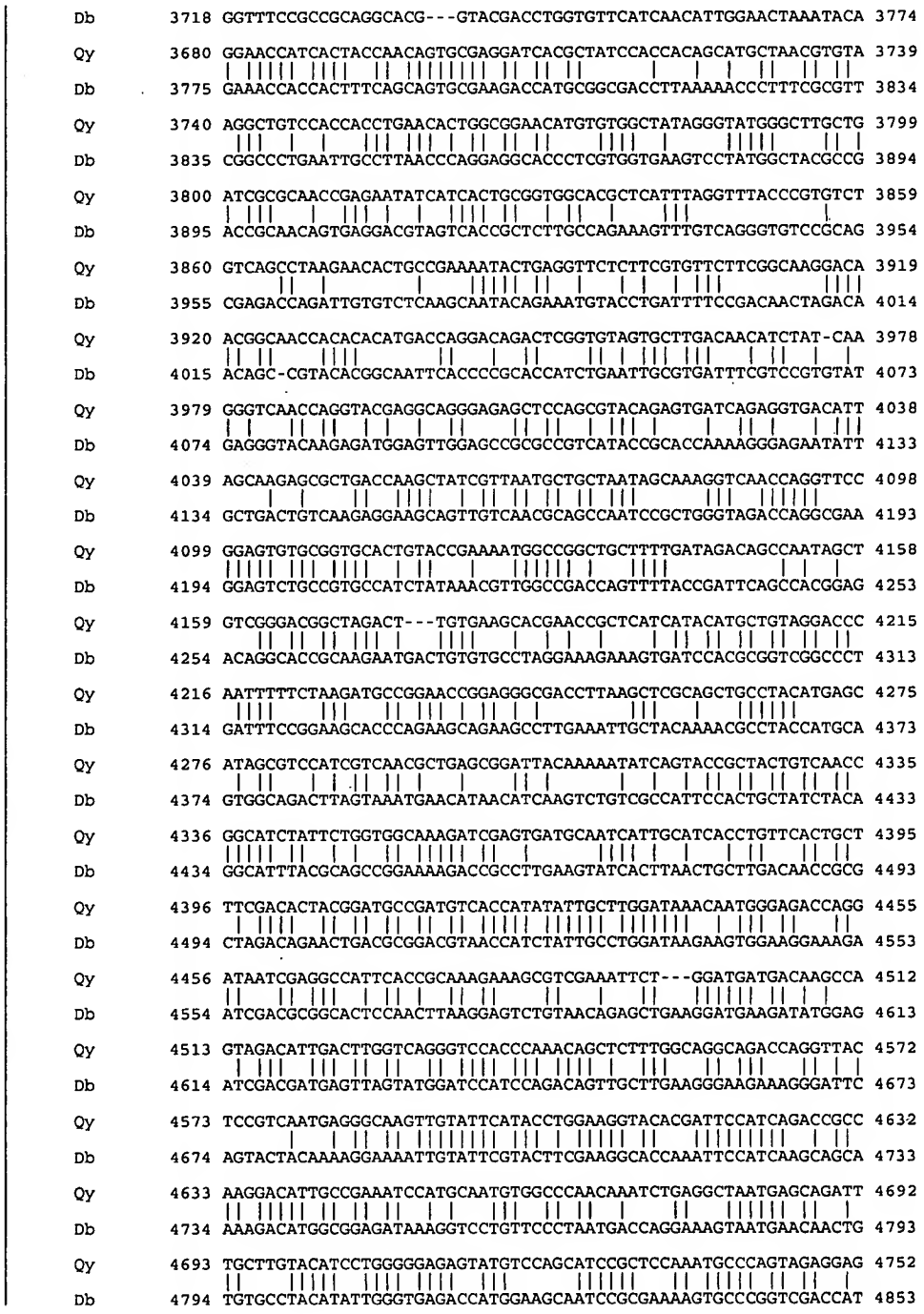
Query Match 26.1%; Score 1303.6; DB 1; Length 11703;
Best Local Similarity 56.1%; Pred. No. 0;
Matches 2821; Conservative 0; Mismatches 2129; Indels 80; Gaps 16;











```
Qy       4753 TCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTACGCTATGACGGCT 4812
              |  |||||   || ||  | ||  | || ||||| || |  || || |||||| |
Db       4854 AACCCGTCGTCTAGCCCGCCCAAAACGTTGCCGTGCCTTTGCATGTATGCCATGACGCCA 4913

Qy       4813 GAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTATGCTCATCATTC 4872
              ||  | ||  |||| |  |        ||      || |  |  | |||||||| ||   |
Db       4914 GAAAGGGTCCACAGACTTAGAAGCAATAACGTCAAAGAAGTTACAGTATGCTCCTCCACC 4973

Qy       4873 CTGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGCAAACCAGTCCTG 4932
              |   | || ||| |||  || |     || ||||||| | |||||||  |||  |||||||
Db       4974 CCCCTTCCTAAGCACAAAATTAAGAATGTTCAGAAGGTTCAGTGCACGAAAGTAGTCCTG 5033

Qy       4933 TTTTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGCGGAAAT 4982
              |||        |   ||  |    || | | || | ||||||   ||| |
Db       5034 TTTAATCCGCACACTCCCGCATTCGTTCCCGCCCGTAAGTACATAGAAGT 5083
```

Search completed: November 15, 2003, 20:02:56
Job time : 382 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 11:11:16 ; Search time 1177 Seconds
(without alignments)
11467.451 Million cell updates/sec

Title: US-10-023-649-1_COPY_1_5000
Perfect score: 5000
Sequence: 1 accctacaaactaatcgatc.....ataattctagaaacgccacc 5000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4492	89.8	11492	20	AAV74107	Western equine enc
2	2118	42.4	11459	21	AAA49442	Venezuelan equine
3	2116.4	42.3	12523	24	AAD29131	Venezuelan equine
4	2113.2	42.3	12379	24	AAD29139	Human immunodefici
5	2113.2	42.3	13584	24	AAD29141	Venezuelan equine
6	2111	42.2	7479	24	AAD29132	Venezuelan equine
7	2074.8	41.5	11464	20	AAV74108	Venezuelan equine
8	1565.6	31.3	11060	20	AAX78130	pSFV link DNA. Se
9	1565.6	31.3	12464	21	AAC55473	Destination vector
10	1564	31.3	7399	25	ACC45139	Alphavirus replica
11	1564	31.3	15538	20	AAX76582	Plasmid pMP44 nucl
12	1562.4	31.2	11489	24	ABN86687	Nucleotide sequenc
13	1562.4	31.2	13599	24	ABN86690	Nucleotide sequenc
14	1554.4	31.1	11517	13	AAQ26021	pSP6-SFV4 RNA tran
15	1473.8	29.5	8010	20	AAX78129	Semliki Forest vir
16	1473.8	29.5	8100	20	AAX76588	SFV EcoRV-SpeI fra
17	1333.2	26.7	11663	19	AAV33473	South African Arbo
18	1333.2	26.7	11663	19	AAV33475	South African Arbo
19	1331.6	26.6	11663	18	AAT47668	South African Arbo
20	1328.4	26.6	11717	21	AAA70577	Sindbis-like virus
21	1306.2	26.1	11626	21	AAA70607	Sindbis-like virus
22	1305.2	26.1	11703	21	AAC64506	Alphavirus SinDCCh
23	1305.2	26.1	11703	21	AAC64507	Alphavirus SinChir
24	1303.6	26.1	11703	19	AAV33476	Sindbis virus clon
25	1302.2	26.0	11717	19	AAV33474	Girdwood S.A.virus
26	1300.4	26.0	8000	25	ABX81525	Sindbis virus cDNA
27	1300.4	26.0	11740	25	ABX81526	Sindbis virus cDNA
28	1300.4	26.0	11927	20	AAX77356	Polynucleotide seq
29	1300.4	26.0	11927	21	AAA90388	Plasmid pSinRep5 h
30	1298.8	26.0	9951	20	AAX77354	Polynucleotide seq
31	1298.8	26.0	9951	21	AAA90386	Sindbis expression
32	1298.8	26.0	9951	22	AAD04742	Alphaviral vector
33	1298.8	26.0	9951	24	ABV73174	Nucleotide sequenc
34	1298.8	26.0	9951	24	ABN86686	Nucleotide sequenc
35	1298.8	26.0	12110	24	ABN86689	Nucleotide sequenc
36	1298.8	26.0	13905	20	AAX77358	Polynucleotide seq
37	1298.8	26.0	13905	21	AAA90390	Plasmid pTE5'2J DN
38	1298.8	26.0	16958	25	AAL55269	DNA of expression
39	1297.2	25.9	10524	20	AAX77355	Polynucleotide seq
40	1297.2	25.9	10524	21	AAA90387	Plasmid pSinRep5 E
41	1295.6	25.9	8000	18	AAX59321	Sindbis virus vari
42	1295.6	25.9	8000	20	AAX58571	Sindbis virus vari
43	1295.6	25.9	8000	25	ABX81524	Sindbis virus cDNA
44	1295.6	25.9	11282	20	AAZ27850	Complete cDNA sequ
45	1295.6	25.9	11282	22	AAF84024	Complete cDNA sequ

ALIGNMENTS

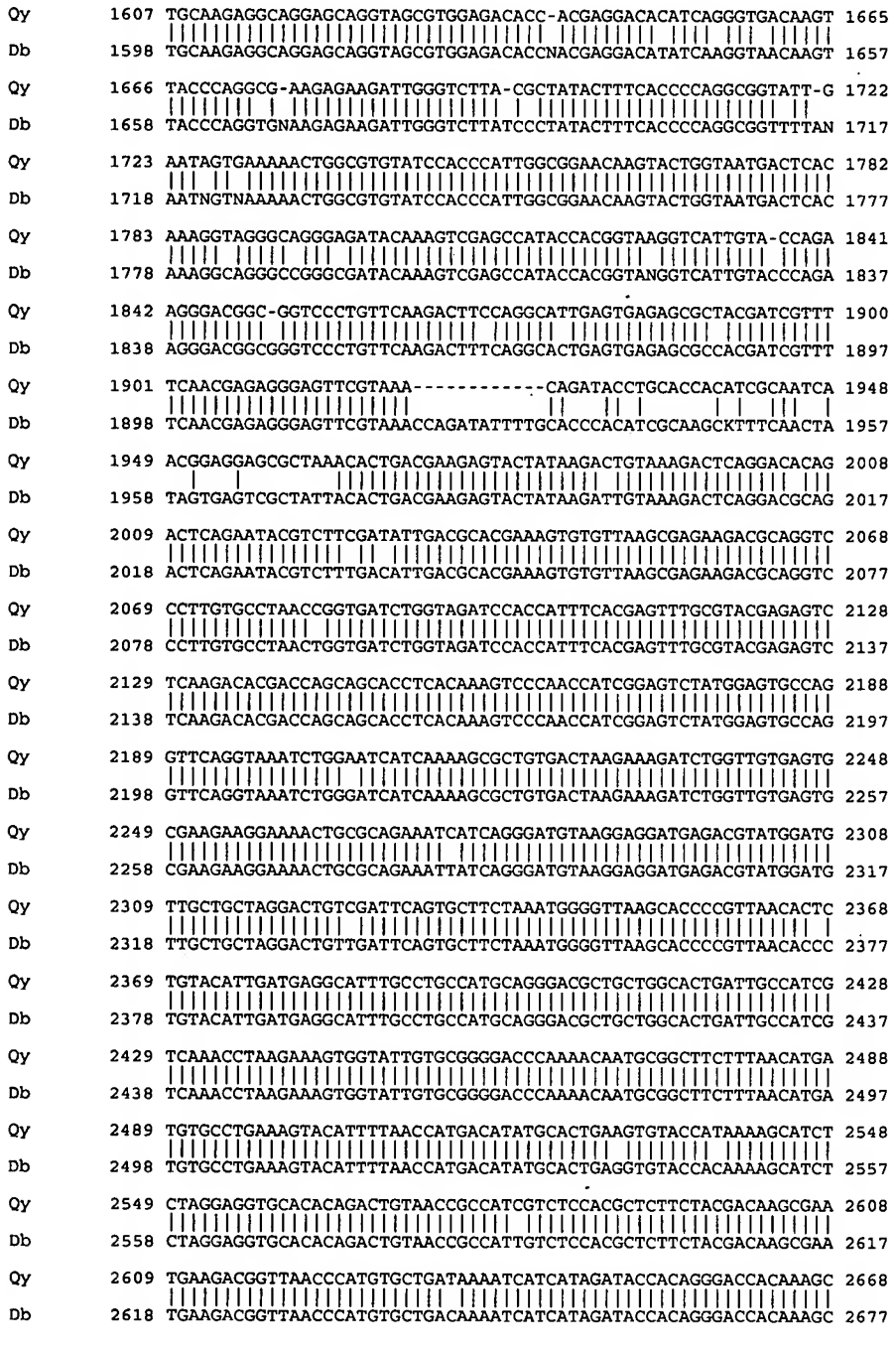
RESULT 1
AAV74107
ID AAV74107 standard; cDNA; 11492 BP.
XX
AC AAV74107;
XX
DT 12-APR-1999 (first entry)
XX
DE Western equine encephalitis virus cDNA.
XX
KW WEE virus; vaccine; ds.
XX
OS Western equine encephalitis virus.
XX
PN WO9853077-A1.
XX
PD 26-NOV-1998.
XX
PF 20-MAY-1998; 98WO-US10645.
XX
PR 16-DEC-1997; 97US-0991840.
PR 20-MAY-1997; 97US-0047162.
PR 24-JUL-1997; 97US-0053652.
XX
PA (REED-) REED ARMY INST RES WALTER.
XX
PI Crise BJ, Oberste MS, Parker MD, Schmura SM, Smith JF;
XX
DR WPI; 1999-045316/04.
XX
PT New DNA encoding infectious Western or Venezuelan equine

PT encephalitis virus genome - useful for the production of live or
PT attenuated vaccines for human or veterinary medicine
XX
PS Example 1; Page 54-67; 112pp; English.
XX
CC This cDNA sequence codes for an infectious western equine
CC encephalitis (WEE) virus RNA transcript. DNA representing the
CC entire genome was prepared by PCR using primers (see AAV74110-21)
CC based on partial genome sequences. 5' Sequences were obtained by
CC RACE. The full-length infectious clone is useful in the production
CC of virulent WEE virus, and for introducing and testing attenuating
CC mutations. Also new are: infectious or attenuated WEE RNA
CC transcripts and WEE viral particles; cDNA (see AAV74108) encoding an
CC infectious Venezuelan equine encephalitis (VEE) variant IE viral
CC genome; infectious or attenuated VEE RNA transcripts and viral
CC particles; an attenuated chimeric virus containing non-structural
CC sequences from a first alpha-virus (aV) and structural sequences
CC from a second aV, resulting in attenuation of the second aV; a method
CC for expressing a protein (especially an antigen to protect against
CC a pathogen) by cloning its gene into an attenuated WEE or VEE IE
CC replicon so that transcription of the replicon produces RNA able to
CC infect the cells in which protein is to be produced; methods for
CC diagnosing WEE and VEE infections; and polypeptides encoded by VEE
CC variant IIIA. Attenuated WEE and VEE are used in live or inactivated
CC vaccines, for use in human or veterinary medicine. Chimeric viruses
CC are also useful as vaccines, directed against the second aV which is
CC particularly Eastern equine encephalitis (EEE) virus or some variant
CC of VEE. WEE and VEE nucleic acids are used as primers and probes to
CC diagnose virus infections and to define natural variants, also for
CC production of protein antigens which can be used as diagnostic
CC reagents, to generate antibodies, and in vaccines. The attenuated
CC viruses are highly immunogenic and provide long-lasting protection.
XX
SQ Sequence 11492 BP; 3257 A; 2854 C; 2794 G; 2570 T; 17 other;

Query Match 89.8%; Score 4492; DB 20; Length 11492;
Best Local Similarity 95.8%; Pred. No. 0;
Matches 4826; Conservative 0; Mismatches 127; Indels 82; Gaps 18;

```
Qy          1 ACCCTACAAACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGC 60
              |||||||||| |||||||||||||||||||||||||||||||||| |||||||||
Db         25 ACCCTACAAACAAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGACGCTGACAGC 84

Qy         61 CCGTATGTCAAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTC 120
              || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         85 CCATATGTCAAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTC 144

Qy        121 ACTGACAATGACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAG 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        145 ACTGACAATGACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAG 204

Qy        181 AGCGAAGTCGACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCA 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        205 AGCGAAGTCGACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCA 264

Qy        241 CATTCCAATCACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGA 300
              |||||||| |||||||||||||||||||||| |||||||||||||||||||||||||||
Db        265 CATTCCAACCACCGCTATCATTGTATCTGCCCCATGATAAGCGCTGAAGACCCGGACAGA 324

Qy        301 CTACAACGGTATGCAGAAAGACTTAAGAAAAGTGACAT-TACCGACAAGAACATAGCCTC 359
              ||||| |||||||||||||||||||||| ||||||||| |||||||||||| ||||||||
Db        325 CTACAGCGGTATGCAGAAAGACTTAAGAGAAGTGACATGTACCGACAAGAATATAGCCTC 384

Qy        360 T-AAGGCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGT 418
              | |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        385 TNAAGGCGGCAGACCTGCTGGAAGTCATGTCCACACCAGACGCAGAGACTCCATCTCTGT 444

Qy        419 GTATGCACACAGACGCCACGTGTAGGTACTTTGGAAGTGTA-GCAGTATACCAAGATGTG 477
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        445 GTATGCACACAGACGCCACGTGTAGGTACTTTGGAAGTGTANGCAGTATACCAAGATGTG 504

Qy        478 TACGCAGTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATT 537
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        505 TACGCAGTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATT 564

Qy        538 TACTGGATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCT 597
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db        565 TACTGGATAGGTTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCT 624

Qy        598 ACTTACAACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGT 657
              |||||||||| |||||||| |||||||||||||||||||||||||||||||||||||||
Db        625 ACTTACAACACAAACTGGGCCGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGT 684

Qy        658 AACTCAGATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTC 717
              |||||||||||||||||||||| |||||||||| |||||||||||||||||||||||||
Db        685 AACTCAGATCTTCAGGAGAGCAGACTTGGAAAACTTTCAATCCTTAGGAAGAAGAGGCTC 744

Qy        718 CAACCTACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCA 777
              |||||||| |||||||||||||||||||||||||||||||||| |||||||| ||||||||
Db        745 CAACCTACTGATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAGGATAGATCA 804

Qy        778 CTGTTACGTAGCTGGCATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACA 837
              |||||||||||||||||||||||||||||||||| |||||||||||| |||||||||||
Db        805 CTGTTACGTAGCTGGCATCTTCCAAACGTGTTCCACCTGAAAGGAAAGTCTNACTTCACA 864

Qy        838 GGTAGATGTGGGACCATTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGC 897
              |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db        865 GGTAGATGTGGGACCATTGTCAGCTGTGAAGGGTATGTCATCAAAAAGATAACGATCAGC 924

Qy        898 CCAGGACTATACGGTAAAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTG 957
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db        925 CCAGGACTATACGGTAAAGTTGAGAACTTGGCGTCCACGATGCATCGCGAGGGTTTCTTG 984

Qy        958 AGTTGCAAAGTCACAGATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTAT 1017
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        985 AGTTGCAAAGTCACAGATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTAT 1044

Qy       1018 GTACCAGCCACACTTTGCGATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGAT 1077
              ||||||||||||||||||||||||||||||||||
Db       1045 GTACCAGCCACACTTTGCGATCAGATGACAGGGA---------------------------- 1078

Qy       1078 GACGCACAAAAACTATTGGTTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAA 1137
                                              |||||||||||| |||||||||||||||||||||||||
Db       1079 ----------------------GGGCTCAACCAACGGATTGTCGTCAATGGTAGGACGCAA 1117

Qy       1138 AGAAATACTAACACAATGCAGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGG 1197
              |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1118 AGAAATACTNACACAATGCAGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGG 1177

Qy       1198 TGGGCGCGTGAACATCGTGCCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGC 1257
              |||||||||| |||||||||||||||||||||||||||| || ||||||||||||||||||
Db       1178 TGGGCGCGTGNACATCGTGCCGACTTGGACGACGAGAAGGAGCTAGGGGTGCGGGAGCGC 1237

Qy       1258 ACTCTTACTATGGGCTGCTGCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAG 1317
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       1238 ACTCTTACTATGGGCTGCTGCTGGGCTTTCAAGACCCAGAAAATTACATCCATCTACAAG 1297

Qy       1318 AAGCCTGGTACGCAAAC-AATTAAGAAAGTACCTGCCGTC-TTTGACTCATTTGTGATTC 1375
              |||||||||||||||| |||||||||||||||||||||| |||||||||||||||||||
Db       1298 AAGCCTGGTACGCAAACAAATTAAGAAAGTACCTGCCGTCTTTTGACTCATTTGTGATTC 1357

Qy       1376 C-ACGCCTTACCAGCCACGGGCTCGAT------ATGGGCTTCCGCCGT-AGGCTCAAGCT 1427
              | |||||||||||||||| |   | |       || |||||||||| ||||||||||
Db       1358 CGACGCCTTACCAGCCACGCGGGGGCTCGAATATGGGCCTTCCGCCGTNAGGCTCAAGCT 1417

Qy       1428 GCTGCTTGAACCAACTGTCAAACCCGCACCGGCTATTACAATGGCCGATGTGGAGCATCT 1487
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db       1418 GCTGCTTGAACCAACTGTCAAACCCGCACCGGCTATTACAATGGCCGATGTGGAGCACCT 1477

Qy       1488 GCGTGGCTTACAGCAAGAAGCTGAAGAAGTGGCTGCAGC-GGAAGAGATCAGAGAAGCCC 1546
              |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db       1478 GCGTGGCTTACAGCAAGAAGCTGAAGAAGTGGCTGCAGCGGGAAGAGATCAGAGAAGCCC 1537

Qy       1547 TGCCACCCTTGCTCCCTGAAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTA 1606
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1538 TGCCACCCTTGCTCCCTGAAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTA 1597
```

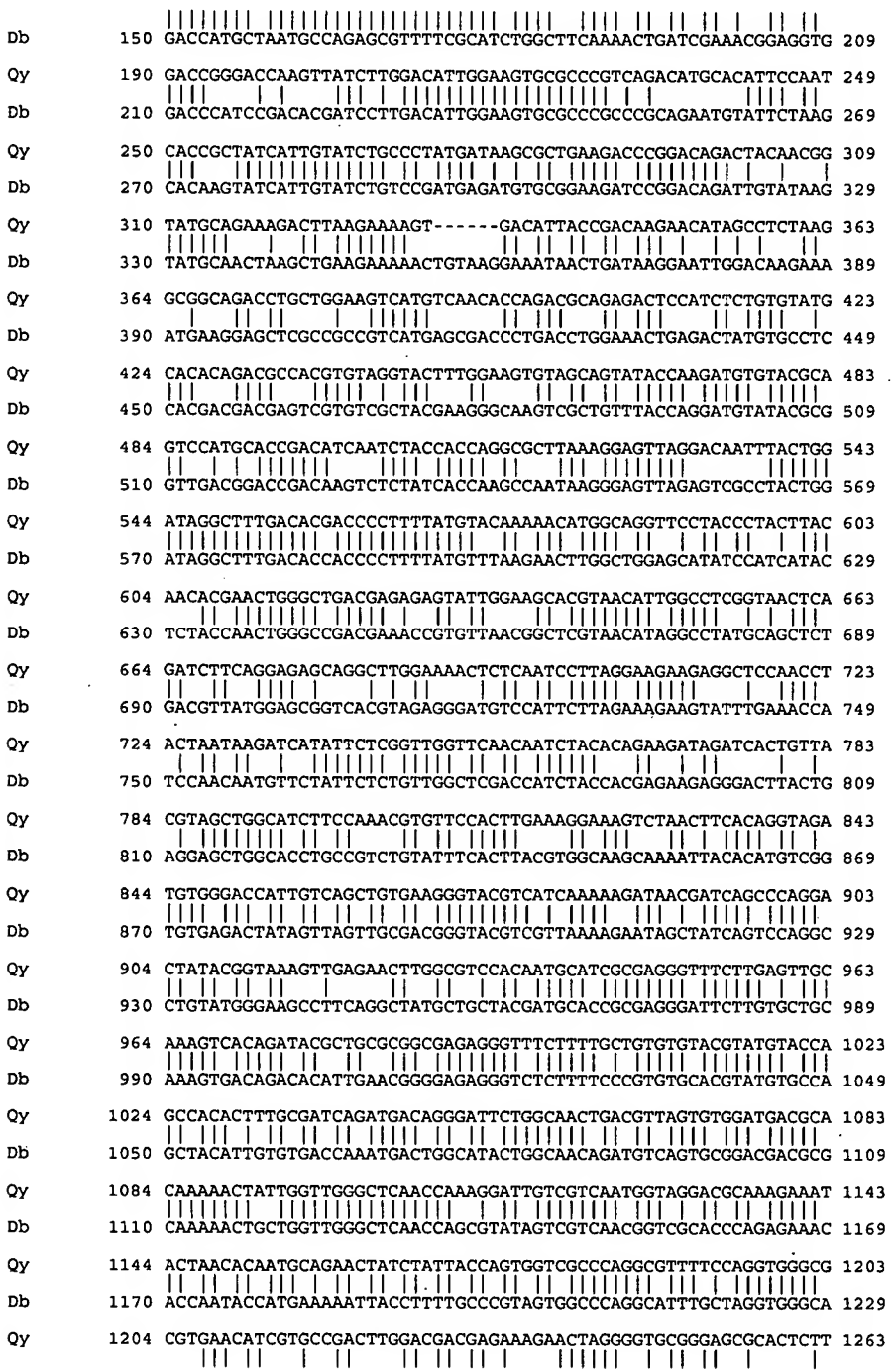

Qy 1607 TGCAAGAGGCAGGAGCAGGTAGCGTGGAGACACC-ACGAGGACACATCAGGGTGACAAGT 1665
Db 1598 TGCAAGAGGCAGGAGCAGGTAGCGTGGAGACACCNACGAGGACATATCAAGGTAACAAGT 1657
Qy 1666 TACCCAGGCG-AAGAGAAGATTGGGTCTTA-CGCTATACTTTCACCCCAGGCGGTATT-G 1722
Db 1658 TACCCAGGTGNAAGAGAAGATTGGGTCTTATCCCTATACTTTCACCCCAGGCGGTTTTAN 1717
Qy 1723 AATAGTGAAAAACTGGCGTGTATCCACCCATTGGCGGAACAAGTACTGGTAATGACTCAC 1782
Db 1718 AATNGTNAAAAACTGGCGTGTATCCACCCATTGGCGGAACAAGTACTGGTAATGACTCAC 1777
Qy 1783 AAAGGTAGGGCAGGGAGATACAAAGTCGAGCCATACCACGGTAAGGTCATTGTA-CCAGA 1841
Db 1778 AAAGGCAGGGCCGGGCGATACAAAGTCGAGCCATACCACGGTANGGTCATTGTACCCAGA 1837
Qy 1842 AGGGACGGC-GGTCCCTGTTCAAGACTTCCAGGCATTGAGTGAGAGCGCTACGATCGTTT 1900
Db 1838 AGGGACGGCGGGTCCCTGTTCAAGACTTTCAGGCACTGAGTGAGAGCGCCACGATCGTTT 1897
Qy 1901 TCAACGAGAGGGAGTTCGTAAA----------CAGATACCTGCACCACATCGCAATCA 1948
Db 1898 TCAACGAGAGGGAGTTCGTAAACCAGATATTTTGCACCCACATCGCAAGCKTTTCAACTA 1957
Qy 1949 ACGGAGGAGCGCTAAACACTGACGAAGAGTACTATAAGACTGTAAAGACTCAGGACACAG 2008
Db 1958 TAGTGAGTCGCTATTACACTGACGAAGAGTACTATAAGATTGTAAAGACTCAGGACGCAG 2017
Qy 2009 ACTCAGAATACGTCTTCGATATTGACGCACGAAAGTGTGTTAAGCGAGAAGACGCAGGTC 2068
Db 2018 ACTCAGAATACGTCTTTGACATTGACGCACGAAAGTGTGTTAAGCGAGAAGACGCAGGTC 2077
Qy 2069 CCTTGTGCCTAACCGGTGATCTGGTAGATCCACCATTTCACGAGTTTGCGTACGAGAGTC 2128
Db 2078 CCTTGTGCCTAACTGGTGATCTGGTAGATCCACCATTTCACGAGTTTGCGTACGAGAGTC 2137
Qy 2129 TCAAGACACGACCAGCAGCACCTCACAAAGTCCCAACCATCGGAGTCTATGGAGTGCCAG 2188
Db 2138 TCAAGACACGACCAGCAGCACCTCACAAAGTCCCAACCATCGGAGTCTATGGAGTGCCAG 2197
Qy 2189 GTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTGACTAAGAAAGATCTGGTTGTGAGTG 2248
Db 2198 GTTCAGGTAAATCTGGGATCATCAAAAGCGCTGTGACTAAGAAAGATCTGGTTGTGAGTG 2257
Qy 2249 CGAAGAAGGAAAACTGCGCAGAAATCATCAGGGATGTAAGGAGGATGAGACGTATGGATG 2308
Db 2258 CGAAGAAGGAAAACTGCGCAGAAATTATCAGGGATGTAAGGAGGATGAGACGTATGGATG 2317
Qy 2309 TTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAATGGGGTTAAGCACCCCGTTAACACTC 2368
Db 2318 TTGCTGCTAGGACTGTTGATTCAGTGCTTCTAAATGGGGTTAAGCACCCCGTTAACACCC 2377
Qy 2369 TGTACATTGATGAGGCATTTGCCTGCCATGCAGGGACGCTGCTGGCACTGATTGCCATCG 2428
Db 2378 TGTACATTGATGAGGCATTTGCCTGCCATGCAGGGACGCTGCTGGCACTGATTGCCATCG 2437
Qy 2429 TCAAACCTAAGAAAGTGGTATTGTGCGGGGACCCAAAACAATGCGGCTTCTTTAACATGA 2488
Db 2438 TCAAACCTAAGAAAGTGGTATTGTGCGGGGACCCAAAACAATGCGGCTTCTTTAACATGA 2497
Qy 2489 TGTGCCTGAAAGTACATTTTAACCATGACATATGCACTGAAGTGTACCATAAAAGCATCT 2548
Db 2498 TGTGCCTGAAAGTACATTTTAACCATGACATATGCACTGAGGTGTACCACAAAAGCATCT 2557
Qy 2549 CTAGGAGGTGCACACAGACTGTAACCGCCATCGTCTCCACGCTCTTCTACGACAAGCGAA 2608
Db 2558 CTAGGAGGTGCACACAGACTGTAACCGCCATTGTCTCCACGCTCTTCTACGACAAGCGAA 2617
Qy 2609 TGAAGACGGTTAACCCATGTGCTGATAAAATCATCATAGATACCACAGGGACCACAAAGC 2668
Db 2618 TGAAGACGGTTAACCCATGTGCTGACAAAATCATCATAGATACCACAGGGACCACAAAGC 2677
Qy 2669 CGCACAAAGATGATCTGATTCTAACCTGTTTCAGAGGATGGGTGAAACAGCTACAGATTG 2728
Db 2678 CGCACAAAGATGATCTGATTCTAACCTGTTTCAGAGGATGGGTGAAACAGCTACAGATTG 2737
Qy 2729 ACTACAAAAATCACGAAATCATGACTGCGGCTGCATCGCAAGGACTTACGCGGAAAGGCG 2788
Db 2738 ACTATAAAAATCATGAAATCATGACTGCGGCTGCATCGCAAGGACTTACGCGGAAAGGCG 2797
Qy 2789 TTTATGCTGTCAGGTACAAAGTCAACGAGAATCCACTCTACTCGCAGACTTCTGAGCACG 2848
Db 2798 TTTATGCTGTCAGGTACAAAGTCAACGAGAATCCACTCTACTCGCAGACTTCTGAGCACG 2857
Qy 2849 TGAACGTGTTACTTACACGCACAGAAAAACGCATTGTCTGGAAGACGCTAGCTGGTGATC 2908
Db 2858 TGAACGTGTTAGTTACACGCACAGAAAAACGCATTGTCTGGAAGACGCTAGCTGGTGACC 2917
Qy 2909 CCTGGATAAAGACACTTACAGCTAAATATCCCGGGGATTTCACGGCTTCATTGGACGACT 2968
Db 2918 CCTGGATAAAGACACTTACAGCTAAATACTCCGGGGATTTCACGGCTTCATTGGACGACT 2977
Qy 2969 GGCAGCGCGAACACGACGCCATTATGGCACGCGTTCTTGATAAGCCGCAGACAGCTGATG 3028
Db 2978 GGCAACGCGAACACGACGCTATTATGGCACGCGTTCTTGATAAGCCGCAGACAGCTGATG 3037
Qy 3029 TGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAGGCTTTAGAGCCAGTCTTGGCCACGG 3088
Db 3038 TGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAGGCTTTAGAGCCAGTCTTGGCCACGG 3097
Qy 3089 CCAACATTGTGCTGACGAGACAGCAGTGGGAGACGTTGCACCCATTCAAGCATGACAGAG 3148
Db 3098 CCAACATTGTGCTGACGAGACAGCAGTGGGAGACGTTGCACCCATTCAAGCATGACAGAG 3157
Qy 3149 CGTACTCACCTGAAATGGCACTGAACTTCTTTTGCACCAGGTTCTTTGGAGTAGACCTGG 3208
Db 3158 CGTACTCACCTGAAATGGCACTGAACTTCTTTTGCACCAGGTTCTTTGGTGTAGACCTGG 3217
Qy 3209 ACAGTGGGTTATTTTCCGCTCCTACCGTCGCACTTACTTACAGGGATCAGCACTGGGATA 3268
Db 3218 ACAGTGGGTTGTTTTCCGCTCCTACCGTCGCACTTACTTACAGGGATCAGCACTGGGACA 3277
Qy 3269 ACTCGCCAGGGAAGAACATGTATGGGCTTAATAGAGAGGTAGCAAAGGAGTTGTCACGGC 3328
Db 3278 ACTCGCCAGGGAAGAACATGTATGGGCTTAATAGAGAGGTAGCAAAGGAGCTGTCACGGC 3337
Qy 3329 GATATCCGTGCATCACAAAAGCGGTTGACACAGGCAGGGTAGCTGATATAAGGAATAATA 3388
Db 3338 GATATCCGTGCATCACAAAAGCGGTTGACACAGGCAGGGTAGCTGATATAAGGAATAATA 3397
Qy 3389 CCATCAAGGACTACTCTCCAACAATTAATGTGGTTCCATTAAATCGCCGGTTGCCCCACT 3448
Db 3398 CCATCAAGGACTACTCTCCAACAATTAATGTGGTTCCATTAAATCGCCGGTTGCCCCACT 3457
Qy 3449 CGTTGATCGTTGACCACAAAGGACAGGGTACAACTGATCACAGCGGATTCCTATCTAAGA 3508
Db 3458 CGCTGATCGTTGACCACAAAGGACAGGGTACAACTGATCACAGTGGATTCCTATCTAAGA 3517
Qy 3509 TGAAGGGCAAATCTGTGTTGGTGATCGGCGATCCTATCAGCATTCCAGGGAAGAAAGTAG 3568
Db 3518 TGAATGGCAAATCTGTGTTGGTGATCGGCGATCCTATCAGCATTCCTGGGAAGAAAGTAG 3577
Qy 3569 AGTCCATGGGTCCATTGCCCACTAATACCATCAGGTGTGATCTCGATTTGGGAATACCTA 3628
Db 3578 AGTCCATGGGTCCATTGCCCACTAATACCATCAGGTGTGATCTCGATTTGGGAATACCTA 3637
Qy 3629 GCCATGTCGGTAAATATGACATTATCTTTGTCAATGTTAGGACCCCGTACAGGAACCATC 3688
Db 3638 GCCATGTCGGTAAATATGACATTATATTTGTCAATGTTAGGACCCCATACAAGAACCATC 3697
Qy 3689 ACTACCAACAGTGCGAGGATCACGCTATCCACCACAGCATGCTAACGTGTAAGGCTGTCC 3748
Db 3698 ACTACCAACAGTGCGAGGATCACGCTATCCACCACAGCATGCTAACGTGTAAGGCTGTCC 3757
Qy 3749 ACCACCTGAACACTGGCGGAACATGTGTGGCTATAGGGTATGGGCTTGCTGATCGCGCAA 3808

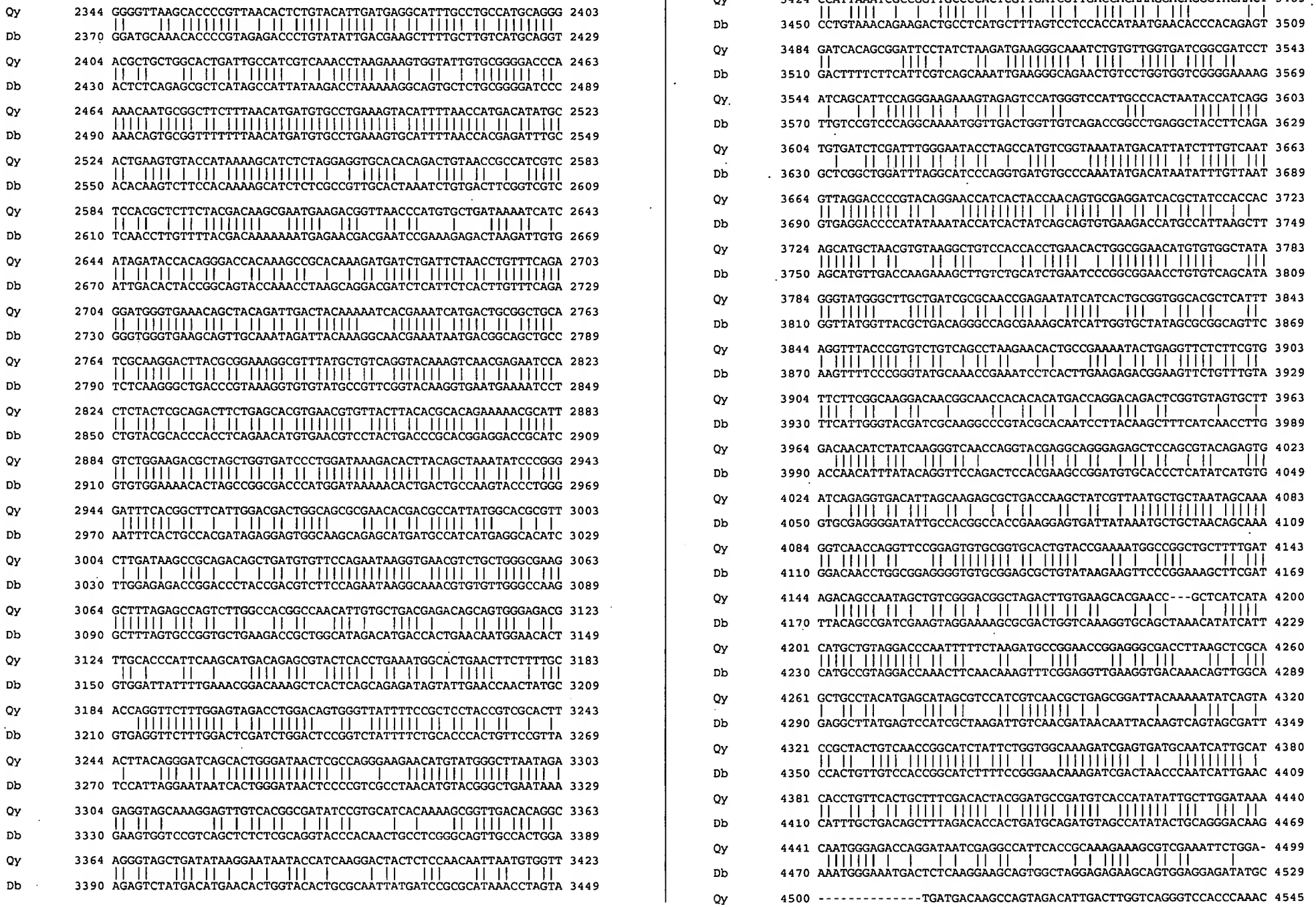
```
Db       3758 ACCACCTGAACACTGGCGGAACATGTGTGGCCATAGGGTATGGACTTGCTGATCGCGCAA 3817

Qy       3809 CCGAGAATATCATCACTGCGGTGGCACGCTCATTTAGGTTTACCCGTGTCTGTCAGCCTA 3868
             |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db       3818 CCGAGAATATCATCACTGCGGTGGCTCGCTCATTTAGGTTTACCCGTGTCTGTCAGCCTA 3877

Qy       3869 AGAACACTGCCGAAAATACTGAGGTTCTCTTCGTGTTCTTCGGCAAGGACAACGGCAACC 3928
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3878 AGAACACTGCCGAAAATACTGAGGTTCTCTTCGTGTTCTTCGGCAAGGACAACGGCAACC 3937

Qy       3929 ACACACATGACCAGGACAGACTCGGTGTAGTGCTTGACAACATCTATCAAGGGTCAACCA 3988
             |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db       3938 ACACACATGACCAGGACAGACTCGGTGTAGTGCTTGACAACATCTACCAAGGGTCAACCA 3997

Qy       3989 GGTACGAGGCAGGGAGAGCTCCAGCGTACAGAGTGATCAGAGGTGACATTAGCAAGAGCG 4048
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3998 GGTACGAGGCAGGGAGAGCTCCAGCGTACAGAGTGATCAGAGGTGACATTAGCAAGAGCG 4057

Qy       4049 CTGACCAAGCTATCGTTAATGCTGCTAATAGCAAAGGTCAACCAGGTTCCGGAGTGTGCG 4108
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4058 CTGACCAAGCTATCGTTAATGCTGCTAATAGCAAAGGTCAACCAGGTTCCGGAGTGTGCG 4117

Qy       4109 GTGCACTGTACCGAAAATGGCCGGCTGCTTTTGATAGACAGCCAATAGCTGTCGGGACGG 4168
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4118 GTGCACTGTACCGAAAATGGCCGGCTGCTTTTGATAGACAGCCAATAGCTGTCGGGACGG 4177

Qy       4169 CTAGACTTGTGAAGCACGAACCGCTCATCATACATGCTGTAGGACCCAATTTTTCTAAGA 4228
             |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db       4178 CTAGACTTGTGAAGCACGAACCGCTCATCATACATGCTGTAGGACCGAATTTTTCTAAGA 4237

Qy       4229 TGCCGGAACCGGAGGGCGACCTTAAGCTCGCAGCTGCCTACATGAGCATAGCGTCCATCG 4288
             |||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4238 TGCCGGAACGAGAGGGCGACCTTAAGCTCGCAGCTGCCTACATGAGCATAGCGTCCATCG 4297

Qy       4289 TCAACGCTGAGCGGATTACAAAAATATCAGTACCGCTACTGTCAACCGGCATCTATTCTG 4348
             |||||||||||||||||||| |||||||||||||||||||||||| ||||||||||||||
Db       4298 TCAACGCTGAGCGGATTACTAAAATATCAGTACCGCTACTGTCAACTGGCATCTATTCTG 4357

Qy       4349 GTGGCAAAGATCGAGTGATGCAATCATTGCATCACCTGTTCACTGCTTTCGACACTACGG 4408
             |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       4358 GTGGCAAAGATCGAGTGATGCAATCATTGCATCACTTGTTCACTGCTTTCGACACTACGG 4417

Qy       4409 ATGCCGATGTCACCATATATTGCTTGGATAAACAATGGGAGACCAGGATAATCGAGGCCA 4468
             |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db       4418 ATGCCGATGTCACCATATATTGCTTGGATAAACAATGGGAGACCAGGATAACCGAGGCCA 4477

Qy       4469 TTCACCGCAAAGAAAGCGTCGAAATTCTG---GATGATGACAAGCCAGTAGACATTGACT 4525
             |||||||||||||||||||||||||| |||   |||||||||||||||||||||||| ||||
Db       4478 TTCACCGCAAAGAAAGCGTCGAAATACTGGATGATGATGACAAGCCAGTAGACATCGACT 4537

Qy       4526 TGGTCAGGGTCCACCCAAACAGCTCTTTGGCAGGCAGACCAGGTTACTCCGTCAATGAGG 4585
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4538 TGGTCAGGGTCCACCCAAACAGCTCTTTGGCAGGCAGACCAGGTTACTCCGTCAATGAGG 4597

Qy       4586 GCAAGTTGTATTCATACCTGGAAGGTACACGATTCCATCAGACCGCCAAGGACATTGCCG 4645
             ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4598 GCAAGCTGTATTCATACCTGGAAGGTACACGATTCCATCAGACCGCCAAGGACATTGCCG 4657

Qy       4646 AAATCCATGCAATGTGGCCCAACAAATCTGAGGCTAATGAGCAGATTTGCTTGTACATCC 4705
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4658 AAATCCATGCAATGTGGCCCAACAAATCTGAGGCTAATGAGCAGATTTGCTTGTACATCC 4717

Qy       4706 TGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGCCCAGTAGAGGAGTCAGAGGCGTCTG 4765
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4718 TGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGCCCAGTAGAGGAGTCAGAGGCGTCTG 4777

Qy       4766 CTCCACCTCACACACTTCCATGCCTGTGTAATTACGCTATGACGGCTGAGCGCGTATACA 4825
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4778 CTCCACCTCACACACTTCCATGCCTGTGTAATTACGCTATGACGGCTGAGCGCGTATACA 4837

Qy       4826 GGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTATGCTCATCATTCCTGTTGCCGAAGT 4885
             |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       4838 GGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTATGCTCATCATTCCCGTTGCCGAAGT 4897

Qy       4886 ACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGCAAACCAGTCCTGTTTTCAGGCGTCG 4945
             |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       4898 ACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGCAAACCAGTCCTGTTCTCAGGCGTCG 4957

Qy       4946 TACCACCGGCTGTACACCCCAGGAAGTACGCGGAAATAATTCTAGAAACGCCACC 5000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4958 TACCACCGGCTGTACACCCCAGGAAGTACGCGGAAATAATTCTAGAAACGCCACC 5012
```

RESULT 2
AAA49442
ID AAA49442 standard; DNA; 11459 BP.
XX
AC AAA49442;
XX
DT 25-SEP-2000 (first entry)
XX
DE Venezuelan equine encephalitis virus (VEE) wild type sequence.
XX
KW Venezuelan equine encephalitis virus; VEE; attenuation; vaccine;
KW replicon; heterologous gene expression; capsid; structural protein;
KW glycoprotein; immunogen; ds.
XX
OS Venezuelan equine encephalitis.
XX
PN WO200034445-A2.
XX
PD 15-JUN-2000.
XX
PF 07-DEC-1999; 99WO-US29041.
XX
PR 07-DEC-1998; 98US-0111330.
XX
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI Pushko P, Parker MD, Smith JF, Crise BJ;
XX
DR WPI; 2000-423405/36.
XX
PT Attenuated Venezuelan equine encephalitis virus useful as a vaccine
PT contains a rearrangement in genome such that its structural
PT glycoproteins precede the capsid gene
XX
PS Disclosure; Page 37-43; 43pp; English.
XX
CC An attenuated Venezuelan equine encephalitis (VEE) has been described
CC in which the virus undergoes a rearrangement of its genome such that
CC the structural glycoproteins precede the capsid gene. The resulting
CC attenuated virus is useful as a live VEE vaccine for human and
CC veterinary use. The attenuated virus can also be used in a replicon
CC expression system. The rearranged VEE replicon can be used to express
CC heterologous genes of interest in cells as well as a means for
CC expressing antigens or immunogenic proteins and peptides of interest.
CC This rearranged sequence appears in the attenuated virus designated
CC RVE1.1.1.
XX
SQ Sequence 11459 BP; 3242 A; 2804 C; 2897 G; 2516 T; 0 other;

Query Match 42.4%; Score 2118; DB 21; Length 11459;
Best Local Similarity 64.5%; Pred. No. 0;
Matches 3219; Conservative 0; Mismatches 1745; Indels 24; Gaps 3;

```
Qy         10 ACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTC 69
              | | | | | |||| ||||| | | |||||||||||| | || | |||||||| |   ||
Db         30 AATTACCTACCCAAAATGGAGAAAGTTCACGTTGACATCGAGGAAGACAGCCCATTCCTC 89

Qy         70 AAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAAT 129
              |   | || ||||||| || || || |||||| | ||||| | |||||||||||| |||
Db         90 AGAGCTTTGCAGCGGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCACTGATAAT 149

Qy        130 GACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTC 189
```


Db 150 GACCATGCTAATGCCAGAGCGTTTTCGCATCTGGCTTCAAAACTGATCGAAACGGAGGTG 209
Qy 190 GACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAAT 249
Db 210 GACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCCGCCCGCAGAATGTATTCTAAG 269
Qy 250 CACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGG 309
Db 270 CACAAGTATCATTGTATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGATTGTATAAG 329
Qy 310 TATGCAGAAAGACTTAAGAAAAGT------GACATTACCGACAAGAACATAGCCTCTAAG 363
Db 330 TATGCAACTAAGCTGAAGAAAAACTGTAAGGAAATAACTGATAAGGAATTGGACAAGAAA 389
Qy 364 GCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATG 423
Db 390 ATGAAGGAGCTCGCCGCCGTCATGAGCGACCCTGACCTGGAAACTGAGACTATGTGCCTC 449
Qy 424 CACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCA 483
Db 450 CACGACGACGAGTCGTGTCGCTACGAAGGGCAAGTCGCTGTTTACCAGGATGTATACGCG 509
Qy 484 GTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGG 543
Db 510 GTTGACGGACCGACAAGTCTCTATCACCAAGCCAATAAGGGAGTTAGAGTCGCCTACTGG 569
Qy 544 ATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTAC 603
Db 570 ATAGGCTTTGACACCACCCCTTTTATGTTTAAGAACTTGGCTGGAGCATATCCATCATAC 629
Qy 604 AACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCA 663
Db 630 TCTACCAACTGGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCCTATGCAGCTCT 689
Qy 664 GATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCT 723
Db 690 GACGTTATGGAGCGGTCACGTAGAGGGATGTCCATTCTTAGAAAGAAGTATTTGAAACCA 749
Qy 724 ACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTA 783
Db 750 TCCAACAATGTTCTATTCTCTGTTGGCTCGACCATCTACCACGAGAAGAGGGACTTACTG 809
Qy 784 CGTAGCTGGCATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGA 843
Db 810 AGGAGCTGGCACCTGCCGTCTGTATTTCACTTACGTGGCAAGCAAAATTACACATGTCGG 869
Qy 844 TGTGGGACCATTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGA 903
Db 870 TGTGAGACTATAGTTAGTTGCGACGGGTACGTCGTTAAAAGAATAGCTATCAGTCCAGGC 929
Qy 904 CTATACGGTAAAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGC 963
Db 930 CTGTATGGGAAGCCTTCAGGCTATGCTGCTACGATGCACCGCGAGGGATTCTTGTGCTGC 989
Qy 964 AAAGTCACAGATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCA 1023
Db 990 AAAGTGACAGACACATTGAACGGGGAGAGGGTCTCTTTTCCCGTGTGCACGTATGTGCCA 1049
Qy 1024 GCCACACTTTGCGATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCA 1083
Db 1050 GCTACATTGTGTGACCAAATGACTGGCATACTGGCAACAGATGTCAGTGCGGACGACGCG 1109
Qy 1084 CAAAAACTATTGGTTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAAT 1143
Db 1110 CAAAAACTGCTGGTTGGGCTCAACCAGCGTATAGTCGTCAACGGTCGCACCCAGAGAAAC 1169
Qy 1144 ACTAACACAATGCAGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCG 1203
Db 1170 ACCAATACCATGAAAAATTACCTTTTGCCCGTAGTGGCCCAGGCATTTGCTAGGTGGGCA 1229
Qy 1204 CGTGAACATCGTGCCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTT 1263
Db 1230 AAGGAATATAAGGAAGATCAAGAAGATGAAAGGCCACTAGGACTACGAGATAGACAGTTA 1289
Qy 1264 ACTATGGGCTGCTGCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCT 1323
Db 1290 GTCATGGGGTGTTGTTGGGCTTTTAGAAGGCACAAGATAACATCTATTTATAAGCGCCCG 1349
Qy 1324 GGTACGCAAACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTT 1383
Db 1350 GATACCCAAACCATCATCAAAGTGAACAGCGATTTCCACTCATTCGTGCTGCCCAGGATA 1409
Qy 1384 ACCAGCCACGGGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACT 1443
Db 1410 GGCAGTAACACATTGGAGATCGGGCTGAGAACAAGAATCAGGAAAATGTTAGAGGAGCAC 1469
Qy 1444 GTCAAACCCGCACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAA 1503
Db 1470 AAGGAGCCGTCACCTCTCATTACCGCCGAGGACGTACAAGAAGCTAAGTGCGCAGCCGAT 1529
Qy 1504 GAAGCTGAAGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCT 1563
Db 1530 GAGGCTAAGGAGGTGCGTGAAGCCGAGGAGTTGCGCGCAGCTCTACCACCTTTGGCAGCT 1589
Qy 1564 GAAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGAGCA 1623
Db 1590 GATGTTGAGGAGCCCACTCTGGAAGCCGATGTCGACTTGATGTTACAAGAGGCTGGGGCC 1649
Qy 1624 GGTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAG 1683
Db 1650 GGCTCAGTGGAGACACCTCGTGGCTTGATAAAGGTTACCAGCTACGCTGGCGAGGACAAG 1709
Qy 1684 ATTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGT 1743
Db 1710 ATCGGCTCTTACGCTGTGCTTTCTCCGCAGGCTGTACTCAAGAGTGAAAAATTATCTTGC 1769
Qy 1744 ATCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATAC 1803
Db 1770 ATCCACCCTCTCGCTGAACAAGTCATAGTGATAACACACTCTGGCCGAAAAGGGCGTTAT 1829
Qy 1804 AAAGTCGAGCCATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAA 1863
Db 1830 GCCGTGGAACCATACCATGGTAAAGTAGTGGTGCCAGAGGGACATGCAATACCCGTCCAG 1889
Qy 1864 GACTTCCAGGCATTGAGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAAC 1923
Db 1890 GACTTTCAAGCTCTGAGTGAAAGTGCCACCATTGTGTACAACGAACGTGAGTTCGTAAAC 1949
Qy 1924 AGATACCTGCACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTAT 1983
Db 1950 AGGTACCTGCACCATATTGCCACACATGGAGGAGCGCTGAACACTGATGAAGAATATTAC 2009
Qy 1984 AAGACTGTAAAGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAG 2043
Db 2010 AAAACTGTCAAGCCCAGCGAGCACGACGGCGAATACCTGTACGACATCGACAGGAAACAG 2069
Qy 2044 TGTGTTAAGCGAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCA 2103
Db 2070 TGCGTCAAGAAAGAACTAGTCACTGGGCTAGGGCTCACAGGCGAGCTGGTGGATCCTCCC 2129
Qy 2104 TTTCACGAGTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCA 2163
Db 2130 TTCCATGAATTCGCCTACGAGAGTCTGAGAACACGACCAGCCGCTCCTTACCAAGTACCA 2189
Qy 2164 ACCATCGGAGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTG 2223
Db 2190 ACCATAGGGGTGTATGGCGTGCCAGGATCAGGCAAGTCTGGCATCATTAAAAGCGCAGTC 2249
Qy 2224 ACTAAGAAAGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGAT 2283
Db 2250 ACCAAAAAAGATCTAGTGGTGAGCGCCAAGAAAGAAAACTGTGCAGAAATTATAAGGGAC 2309
Qy 2284 GTAAGGAGGATGAGACGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAAT 2343
Db 2310 GTCAAGAAAATGAAAGGGCTGGACGTCAATGCCAGAACTGTGGACTCAGTGCTCTTGAAT 2369

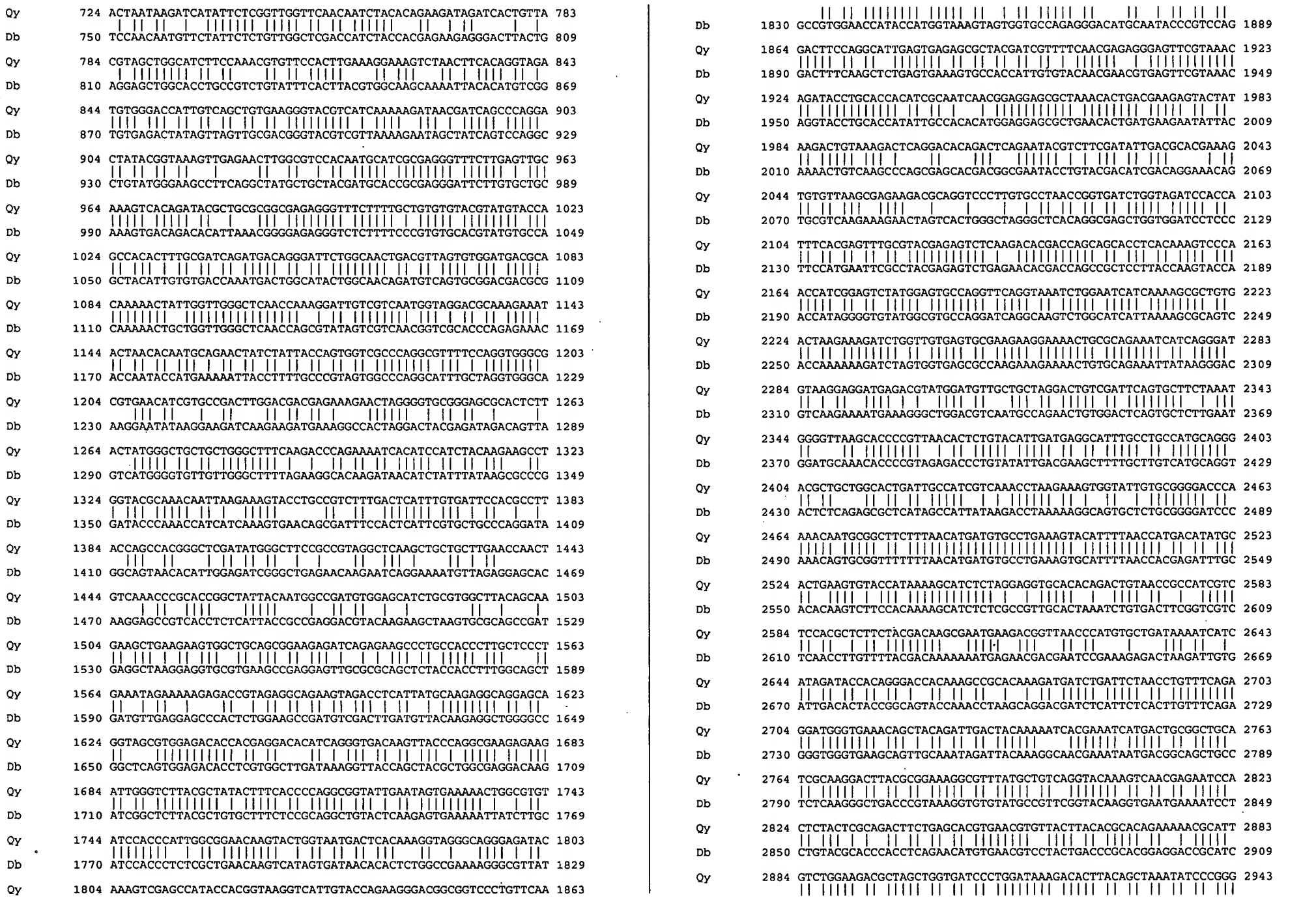


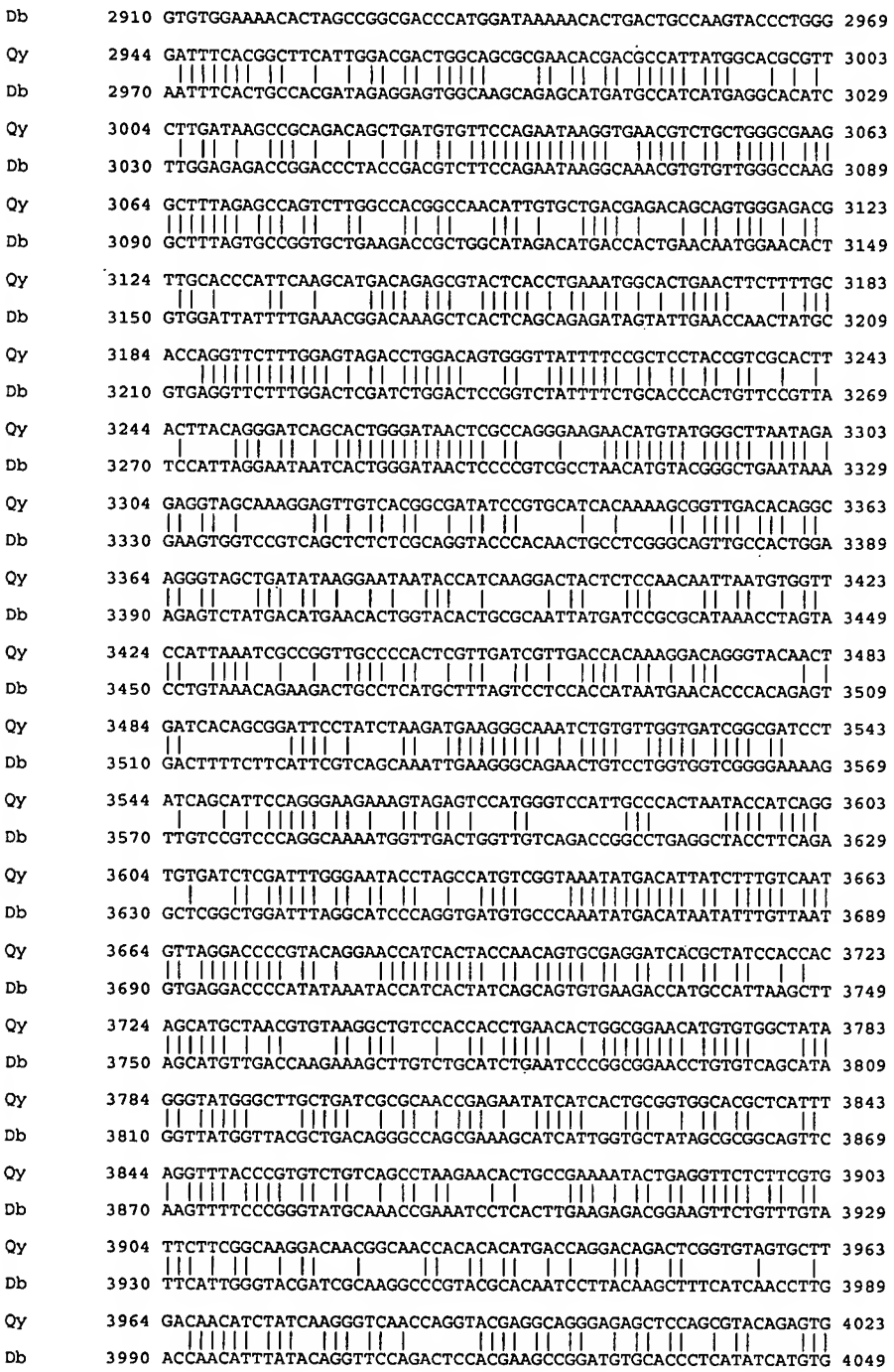
```
Db       4530 ATATCCGACGACTCTTCAGTGACAGAACCTGATGCAGAGCTGGTGAGGGTGCATCCGAAG 4589

Qy       4546 AGCTCTTTGGCAGGCAGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTG 4605

Db       4590 AGTTCTTTGGCTGGAAGGAAGGGCTACAGCACAAGCGATGGCAAAACTTTCTCATATTTG 4649

Qy       4606 GAAGGTACACGATTCCATCAGACCGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCC 4665

Db       4650 GAAGGGACCAAGTTTCACCAGGCGGCCAAGGATATAGCAGAAATTAATGCCATGTGGCCC 4709

Qy       4666 AACAAATCTGAGGCTAATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGC 4725

Db       4710 GTTGCAACGGAGGCCAATGAGCAGGTATGCATGTATATCCTCGGAGAAAGCATGAGCAGT 4769

Qy       4726 ATCCGCTCCAAATGCCCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCA 4785

Db       4770 ATTAGGTCGAAATGCCCCGTCGAAGAGTCGGAAGCCTCCACACCACCTAGCACGCTGCCT 4829

Qy       4786 TGCCTGTGTAATTACGCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAA 4845

Db       4830 TGCTTGTGCATCCATGCCATGACTCCAGAAAGAGTACAGCGCCTAAAAGCCTCACGTCCA 4889

Qy       4846 GAACAGTTCGCCGTATGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAG 4905

Db       4890 GAACAAATTACTGTGTGCTCATCCTTTCCATTGCCGAAGTATAGAATCACTGGTGTGCAG 4949

Qy       4906 AAGCTACAGTGCAGCAAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCC 4965

Db       4950 AAGATCCAATGCTCCCAGCCTATATTGTTCTCACCGAAAGTGCCTGCGTATATTCATCCA 5009

Qy       4966 AGGAAGTA 4973

Db       5010 AGGAAGTA 5017
```

RESULT 3
AAD29131
ID AAD29131 standard; DNA; 12523 BP.
XX
AC AAD29131;
XX
DT 07-MAY-2002 (first entry)
XX
DE Venezuelan equine encephalitis virus (VEE) replicon plasmid.
XX
KW Venezuelan equine encephalitis virus; VEE; vaccine; HIV infection;
KW human immunodeficiency virus; alpha-virus replicon; immune response;
KW therapy; ds.
XX
OS Chimeric - Venezuelan equine encephalitis virus.
OS Chimeric - Unidentified.
XX
PN WO200203917-A2.
XX
PD 17-JAN-2002.
XX
PF 09-JUL-2001; 2001WO-US21701.
XX
PR 07-JUL-2000; 2000US-216995P.
XX
PA (ALPH-) ALPHAVAX INC.
PA (UYNC-) UNIV NORTH CAROLINA.
XX
PI Olmsted R, Keith P, Dryga S, Caley I, Maughan M, Johnston R;
PI Davis N, Swanstrom R;
XX
DR WPI; 2002-171664/22.
XX
PT Composition useful for treating or preventing HIV infections, comprises
PT two or more isolated nucleic acids encoding env, gag or pol gene
PT product of HIV or immunogenic fragment of the gene products -
XX
PS Example 5; Page 141-145; 201pp; English.
XX
CC The invention relates to a composition comprising isolated nucleic acids
CC encoding env, gag or pol gene product of human immunodeficiency virus or
CC immunogenic fragment of the gene products. The gag gene product is
CC modified to inhibit formation of virus-like particles containing gag
CC gene product and their release from cells, and the pol gene product is
CC modified to inhibit reverse transcriptase activity. The invention also
CC relates to a method for producing an alpha-virus replicon particle used
CC in vaccines. The composition is useful for inducing an immune response
CC to human immunodeficiency virus (HIV) or for treating or preventing HIV
CC infection in a subject. The alpha-virus replicon particle is useful in a
CC vaccine. The composition is useful for administering a protein or peptide
CC to a subject. A composition comprising heparin affinity-purified alpha-
CC virus replicon particle is useful as a clinical trial material and as a
CC commercial product. The present sequence is Venezuelan equine
CC encephalitis virus (VEE) replicon plasmid containing the DU422 gag gene
CC (p3-40.1.6) used in the exemplification of the invention.
XX
SQ Sequence 12523 BP; 3579 A; 2959 C; 3160 G; 2825 T; 0 other;

Query Match 42.3%; Score 2116.4; DB 24; Length 12523;
Best Local Similarity 64.5%; Pred. No. 0;
Matches 3218; Conservative 0; Mismatches 1746; Indels 24; Gaps 3;

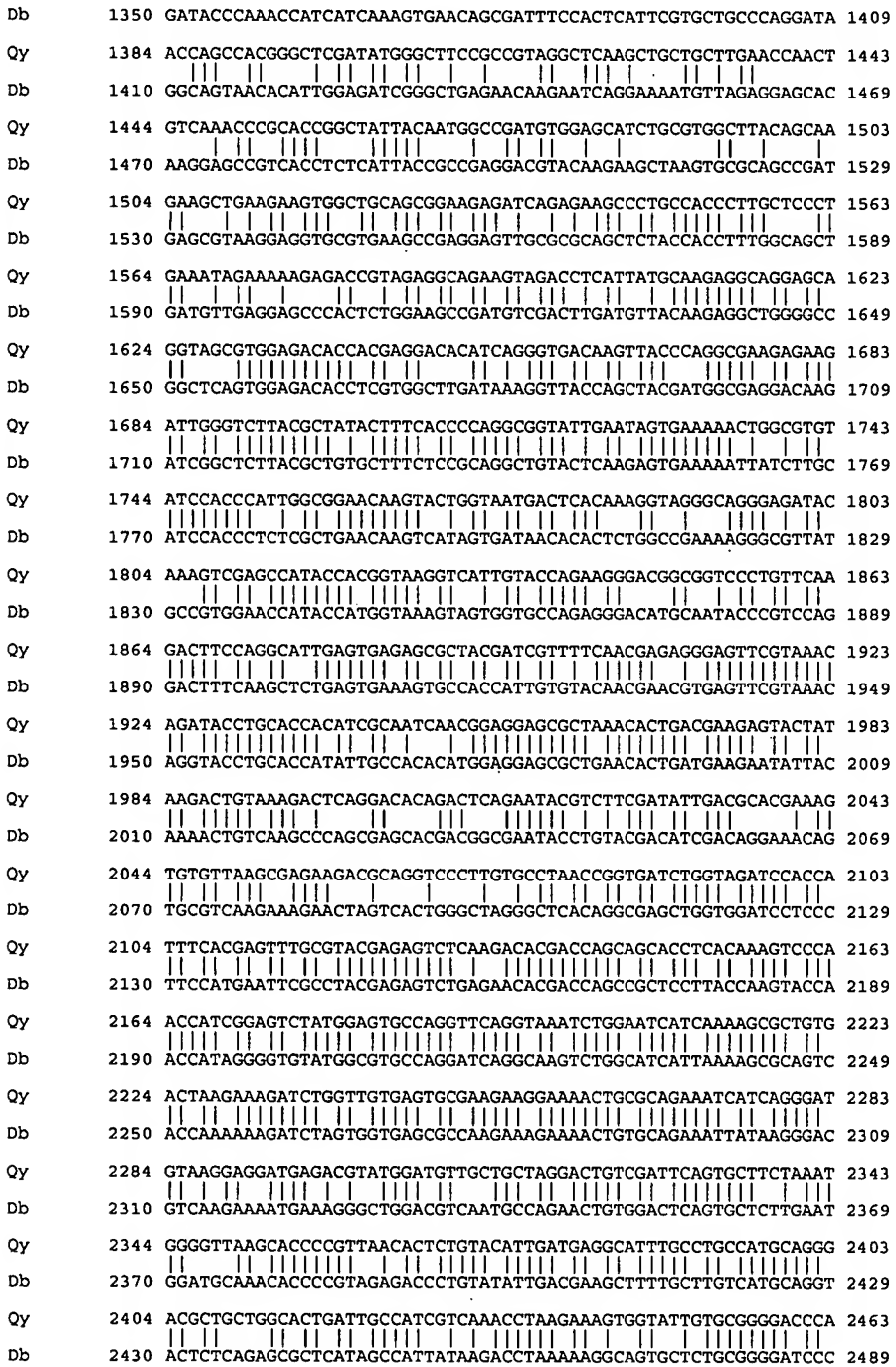
```
Qy         10 ACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTC 69

Db         30 AATTACCTACCCAAAATGGAGAAAGTTCACGTTGACATCGAGGAAGACAGCCCATTCCTC 89

Qy         70 AAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAAT 129

Db         90 AGAGCTTTGCAGCGGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCACTGATAAT 149

Qy        130 GACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTC 189

Db        150 GACCATGCTAATGCCAGAGCGTTTTCGCATCTGGCTTCAAAACTGATCGAAACGGAGGTG 209

Qy        190 GACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAAT 249

Db        210 GACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCCGCCCGCAGAATGTATTCTAAG 269

Qy        250 CACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGG 309

Db        270 CACAAGTATCATTGTATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGATTGTATAAG 329

Qy        310 TATGCAGAAAGACTTAAGAAAAGT------GACATTACCGACAAGAACATAGCCTCTAAG 363

Db        330 TATGCAACTAAGCTGAAGAAAAACTGTAAGGAAATAACTGATAAGGAATTGGACAAGAAA 389

Qy        364 GCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATG 423

Db        390 ATGAAGGAGCTCGCCGCCGTCATGAGCGACCCTGACCTGGAAACTGAGACTATGTGCCTC 449

Qy        424 CACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCA 483

Db        450 CACGACGACGAGTCGTGTCGCTACGAAGGGCAAGTCGCTGTTTACCAGGATGTATACGCG 509

Qy        484 GTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGG 543

Db        510 GTTGACGGACCGACAAGTCTCTATCACCAAGCCAATAAGGGAGTTAGAGTCGCCTACTGG 569

Qy        544 ATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTAC 603

Db        570 ATAGGCTTTGACACCACCCCTTTTATGTTTAAGAACTTGGCTGGAGCATATCCATCATAC 629

Qy        604 AACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCA 663

Db        630 TCTACCAACTGGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCCTATGCAGCTCT 689

Qy        664 GATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCT 723

Db        690 GACGTTATGGAGCGGTCACGTAGAGGGATGTCCATTCTTAGAAAGAAGTATTTGAAACCA 749
```





RESULT 4
AAD29139
ID AAD29139 standard; DNA; 12379 BP.

```
XX
AC   AAD29139;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Human immunodeficiency virus type 1 (HIV-1) pol plasmid.
XX
KW   Human immunodeficiency virus; vaccine; HIV infection; immune response;
KW   alpha-virus replicon; therapy; pol gene; ds.
XX
OS   Chimeric - Human immunodeficiency virus type 1.
OS   Chimeric - Unidentified.
XX
PN   WO200203917-A2.
XX
PD   17-JAN-2002.
XX
PF   09-JUL-2001; 2001WO-US21701.
XX
PR   07-JUL-2000; 2000US-216995P.
XX
PA   (ALPH-) ALPHAVAX INC.
PA   (UYNC-) UNIV NORTH CAROLINA.
XX
PI   Olmsted R,  Keith P,  Dryga S,  Caley I,  Maughan M,  Johnston R;
PI   Davis N,  Swanstrom R;
XX
DR   WPI; 2002-171664/22.
XX
PT   Composition useful for treating or preventing HIV infections, comprises
PT   two or more isolated nucleic acids encoding env, gag or pol gene
PT   product of HIV or immunogenic fragment of the gene products  -
XX
PS   Example 5; Page 183-187; 201pp; English.
XX
CC   The invention relates to a composition comprising isolated nucleic acids
CC   encoding env, gag or pol gene product of human immunodeficiency virus or
CC   immunogenic fragment of the gene products. The gag gene product is
CC   modified to inhibit formation of virus-like particles containing gag
CC   gene product and their release from cells, and the pol gene product is
CC   modified to inhibit reverse transcriptase activity. The invention also
CC   relates to a method for producing an alpha-virus replicon particle used
CC   in vaccines. The composition is useful for inducing an immune response
CC   to human immunodeficiency virus (HIV) or for treating or preventing HIV
CC   infection in a subject. The alpha-virus replicon particle is useful in a
CC   vaccine. The composition is useful for administering a protein or peptide
CC   to a subject. A composition comprising heparin affinity-purified alpha-
CC   virus replicon particle is useful as a clinical trial material and as a
CC   commercial product. The present sequence is HIV-1 pol plasmid.
XX
SQ   Sequence 12379 BP; 3562 A; 2920 C; 3084 G; 2813 T; 0 other;

  Query Match             42.3%;  Score 2113.2;  DB 24;  Length 12379;
  Best Local Similarity   64.5%;  Pred. No. 0;
  Matches 3216;  Conservative    0;  Mismatches 1748;  Indels   24;  Gaps    3;

Qy         10 ACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTC 69
              | | | | | |||| ||||| | | |||||||||||| | || |  |||||||| |   ||
Db         30 AATTACCTACCCAAAATGGAGAAAGTTCACGTTGACATCGAGGAAGACAGCCCATTCCTC 89

Qy         70 AAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAAT 129
              |   | || ||||||| || || || |||||| | ||||| | |||||||||||| |||
Db         90 AGAGCTTTGCAGCGGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCACTGATAAT 149

Qy        130 GACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTC 189
              |||||||| |||||||||||||||||||||| ||||  |||| || || || |  || ||
Db        150 GACCATGCTAATGCCAGAGCGTTTTCGCATCTGGCTTCAAAACTGATCGAAACGGAGGTG 209

Qy        190 GACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAAT 249
              ||||    | |    ||| | ||||||||||||||||||| | |         |||| ||
Db        210 GACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCCGCCCGCAGAATGTATTCTAAG 269

Qy        250 CACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGG 309
              ||| |||||  || |  |||||      ||  ||  ||||||| ||| | ||  |  |
Db        270 CACAAGTATCATTGTATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGATTGTATAAG 329

Qy        310 TATGCAGAAAGACTTAAGAAAAGT------GACATTACCGACAAGAACATAGCCTCTAAG 363
              ||||||    |  || |||||||         || || || || ||| |  | | |   ||
Db        330 TATGCAACTAAGCTGAAGAAAAACTGTAAGGAAATAACTGATAAGGAATTGGACAAGAAA 389

Qy        364 GCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATG 423
                 |   || ||    |  ||||||      || |||   || |||    || ||||  |
Db        390 ATGAAGGAGCTCGCCGCCGTCATGAGCGACCCTGACCTGGAAACTGAGACTATGTGCCTC 449

Qy        424 CACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCA 483
              |||   ||||   ||||| | |||    ||    || || || ||||| ||||| |||||
Db        450 CACGACGACGAGTCGTGTCGCTACGAAGGGCAAGTCGCTGTTTACCAGGATGTATACGCG 509

Qy        484 GTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGG 543
              ||  | | |||||||     |||| |||||  ||   ||| ||||||||        ||||||
Db        510 GTTGACGGACCGACAAGTCTCTATCACCAAGCCAATAAGGGAGTTAGAGTCGCCTACTGG 569

Qy        544 ATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTAC 603
              |||||||||||||| ||||||||||||  || ||| |||| ||  | || ||  | |||
Db        570 ATAGGCTTTGACACCACCCCTTTTATGTTTAAGAACTTGGCTGGAGCATATCCATCATAC 629

Qy        604 AACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCA 663
                 || |||||||| ||||| |  || ||    || ||||||| |||||  | | |||
Db        630 TCTACCAACTGGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCCTATGCAGCTCT 689

Qy        664 GATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCT 723
              ||  ||  |||| |     | | ||     | || || ||||| ||||||      |  ||||
Db        690 GACGTTATGGAGCGGTCACGTAGAGGGATGTCCATTCTTAGAAAGAAGTATTTGAAACCA 749

Qy        724 ACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTA 783
              | || ||  |  ||||||| ||||| || || ||||||   ||  | ||      |  |
Db        750 TCCAACAATGTTCTATTCTCTGTTGGCTCGACCATCTACCACGAGAAGAGGGACTTACTG 809

Qy        784 CGTAGCTGGCATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGA 843
              | |||||||| || ||     || || |||||     || |||   || | |||| || |
Db        810 AGGAGCTGGCACCTGCCGTCTGTATTTCACTTACGTGGCAAGCAAAATTACACATGTCGG 869

Qy        844 TGTGGGACCATTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGA 903
              |||| ||| || || || || || ||||||||| | ||||  ||| | |||||  |||||
Db        870 TGTGAGACTATAGTTAGTTGCGACGGGTACGTCGTTAAAAGAATAGCTATCAGTCCAGGC 929

Qy        904 CTATACGGTAAAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGC 963
              || || || ||    |     ||  ||  | || ||||| |||||||| |||||| | |||
Db        930 CTGTATGGGAAGCCTTCAGGCTATGCTGCTACGATGCACCGCGAGGGATTCTTGTGCTGC 989

Qy        964 AAAGTCACAGATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCA 1023
              ||||| ||||| ||  |   ||| |||||||| |||||| | ||||| |||||||| |||
Db        990 AAAGTGACAGACACATTCAACGGGGAGAGGGTCTCTTTTCCCGTGTGCACGTATGTGCCA 1049

Qy       1024 GCCACACTTTGCGATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCA 1083
              || ||| | || || || ||||| || || |||||||| || || |||| ||| |||||
Db       1050 GCTACATTGTGTGACCAAATGACTGGCATACTGGCAACAGATGTCAGTGCGGACGACGCG 1109

Qy       1084 CAAAAACTATTGGTTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAAT 1143
              |||||||| |||||||||||||||  | || |||||||| ||| | || || |||||
Db       1110 CAAAAACTGCTGGTTGGGCTCAACCAGCGTATAGTCGTCAACGGTCGCACCCAGAGAAAC 1169

Qy       1144 ACTAACACAATGCAGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCG 1203
              || || || ||| | || || || || || || || |||||||| ||| | ||||||||
Db       1170 ACCAATACCATGAAAAATTACCTTTGCCCGTAGTGGCCCAGGCATTTGCTAGGTGGGCA 1229

Qy       1204 CGTGAACATCGTGCCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTT 1263
              ||| ||   |  ||   || || || |    ||||||  | || ||  |      |
Db       1230 AAGGAATATAAGGAAGATCAAGAAGATGAAAGGCCACTAGGACTACGAGATAGACAGTTA 1289

Qy       1264 ACTATGGGCTGCTGCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCT 1323
              |||||  || || |||||||| |  |  || || || ||||| || || |||   ||
Db       1290 GTCATGGGGTGTTGTTGGGCTTTTAGAAGGCACAAGATAACATCTATTTATAAGCGCCCG 1349

Qy       1324 GGTACGCAAACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTT 1383
              | ||| |||||  || |  |||||      ||  ||  ||||||| ||| | ||  |  |
```

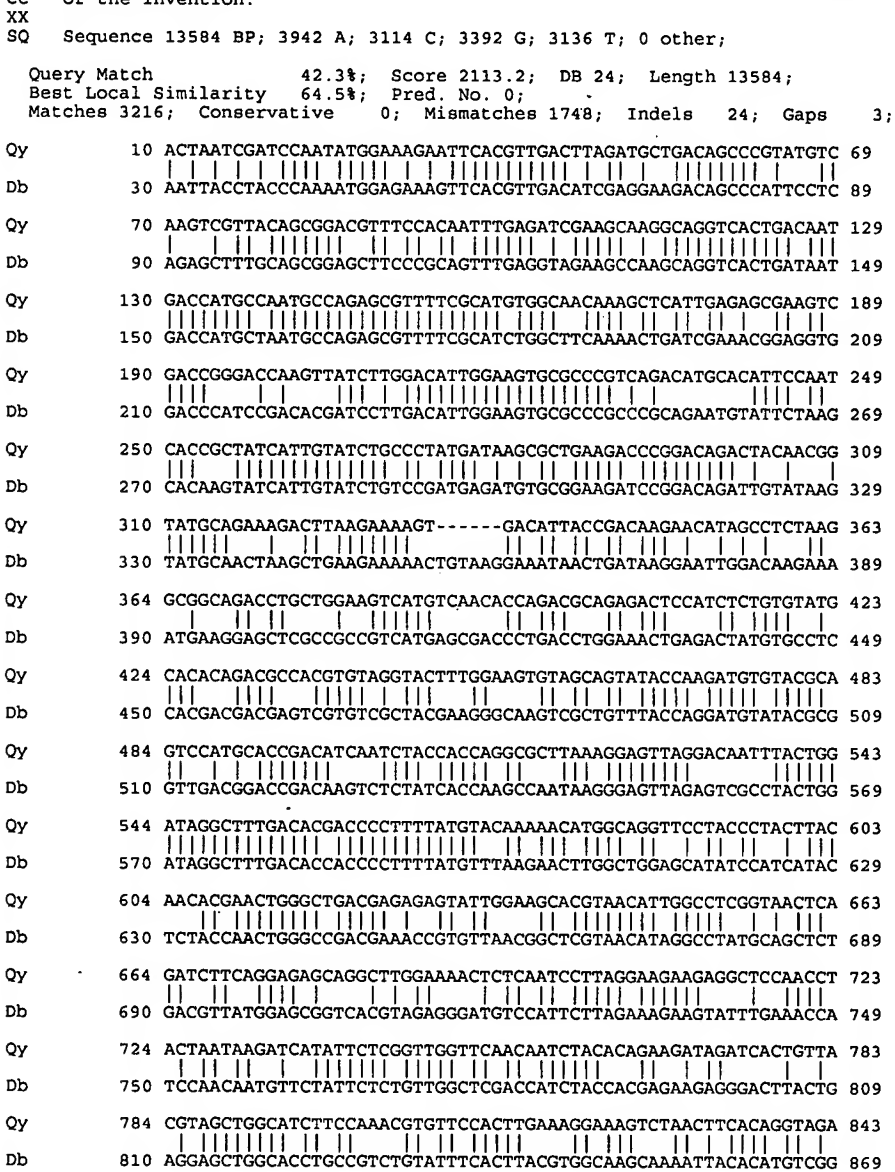

Db 1350 GATACCCAAACCATCATCAAAGTGAACAGCGATTTCCACTCATTCGTGCTGCCCAGGATA 1409
Qy 1384 ACCAGCCACGGGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACT 1443
Db 1410 GGCAGTAACACATTGGAGATCGGGCTGAGAACAAGAATCAGGAAAATGTTAGAGGAGCAC 1469
Qy 1444 GTCAAACCCGCACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAA 1503
Db 1470 AAGGAGCCGTCACCTCTCATTACCGCCGAGGACGTACAAGAAGCTAAGTGCGCAGCCGAT 1529
Qy 1504 GAAGCTGAAGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCT 1563
Db 1530 GAGCGTAAGGAGGTGCGTGAAGCCGAGGAGTTGCGCGCAGCTCTACCACCTTTGGCAGCT 1589
Qy 1564 GAAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAGGCAGGAGCA 1623
Db 1590 GATGTTGAGGAGCCCACTCTGGAAGCCGATGTCGACTTGATGTTACAAGAGGCTGGGGCC 1649
Qy 1624 GGTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAG 1683
Db 1650 GGCTCAGTGGAGACACCTCGTGGCTTGATAAAGGTTACCAGCTACGATGGCGAGGACAAG 1709
Qy 1684 ATTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGT 1743
Db 1710 ATCGGCTCTTACGCTGTGCTTTCTCCGCAGGCTGTACTCAAGAGTGAAAAATTATCTTGC 1769
Qy 1744 ATCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATAC 1803
Db 1770 ATCCACCCTCTCGCTGAACAAGTCATAGTGATAACACACTCTGGCCGAAAAGGGCGTTAT 1829
Qy 1804 AAAGTCGAGCCATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAA 1863
Db 1830 GCCGTGGAACCATACCATGGTAAAGTAGTGGTGCCAGAGGGACATGCAATACCCGTCCAG 1889
Qy 1864 GACTTCCAGGCATTGAGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAAC 1923
Db 1890 GACTTTCAAGCTCTGAGTGAAAGTGCCACCATTGTGTACAACGAACGTGAGTTCGTAAAC 1949
Qy 1924 AGATACCTGCACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTAT 1983
Db 1950 AGGTACCTGCACCATATTGCCACACATGGAGGAGCGCTGAACACTGATGAAGAATATTAC 2009
Qy 1984 AAGACTGTAAAGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAG 2043
Db 2010 AAAACTGTCAAGCCCAGCGAGCACGACGGCGAATACCTGTACGACATCGACAGGAAACAG 2069
Qy 2044 TGTGTTAAGCGAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCA 2103
Db 2070 TGCGTCAAGAAAGAACTAGTCACTGGGCTAGGGCTCACAGGCGAGCTGGTGGATCCTCCC 2129
Qy 2104 TTTCACGAGTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCA 2163
Db 2130 TTCCATGAATTCGCCTACGAGAGTCTGAGAACACGACCAGCCGCTCCTTACCAAGTACCA 2189
Qy 2164 ACCATCGGAGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTG 2223
Db 2190 ACCATAGGGGTGTATGGCGTGCCAGGATCAGGCAAGTCTGGCATCATTAAAAGCGCAGTC 2249
Qy 2224 ACTAAGAAAGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGAT 2283
Db 2250 ACCAAAAAAAGATCTAGTGGTGAGCGCCAAGAAAGAAAACTGTGCAGAAATTATAAGGGAC 2309
Qy 2284 GTAAGGAGGATGAGACGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAAT 2343
Db 2310 GTCAAGAAAATGAAAGGGCTGGACGTCAATGCCAGAACTGTGGACTCAGTGCTCTTGAAT 2369
Qy 2344 GGGGTTAAGCACCCCGTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGG 2403
Db 2370 GGATGCAAACACCCCGTAGAGACCCTGTATATTGACGAAGCTTTTGCTTGTCATGCAGGT 2429
Qy 2404 ACGCTGCTGGCACTGATTGCCATCGTCAAACCTAAGAAAGTGGTATTGTGCGGGGACCCA 2463
Db 2430 ACTCTCAGAGCGCTCATAGCCATTATAAGACCTAAAAAGGCAGTGCTCTGCGGGGATCCC 2489
Qy 2464 AAACAATGCGGCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGC 2523
Db 2490 AAACAGTGCGGTTTTTTTAACATGATGTGCCTGAAAGTGCATTTTAACCACGAGATTTGC 2549
Qy 2524 ACTGAAGTGTACCATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATCGTC 2583
Db 2550 ACACAAGTCTTCCACAAAAGCATCTCTCGCCGTTGCACTAAATCTGTGACTTCGGTCGTC 2609
Qy 2584 TCCACGCTCTTCTACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATCATC 2643
Db 2610 TCAACCTTGTTTTACGACAAAAAAATGAGAACGACGAATCCGAAAGAGACTAAGATTGTG 2669
Qy 2644 ATAGATACCACAGGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGA 2703
Db 2670 ATTGACACTACCGGCAGTACCAAACCTAAGCAGGACGATCTCATTCTCACTTGTTTCAGA 2729
Qy 2704 GGATGGGTGAAACAGCTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCTGCA 2763
Db 2730 GGGTGGGTGAAGCAGTTGCAAATAGATTACAAAGGCAACGAAATAATGACGGCAGCTGCC 2789
Qy 2764 TCGCAAGGACTTACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCA 2823
Db 2790 TCTCAAGGGCTGACCCGTAAAGGTGTGTATGCCGTTCGGTACAAGGTGAATGAAAATCCT 2849
Qy 2824 CTCTACTCGCAGACTTCTGAGCACGTGAACGTGTTACTTACACGCACAGAAAAACGCATT 2883
Db 2850 CTGTACGCACCCACCTCAGAACATGTGAACGTCCTACTGACCCGCACGGAGGACCGCATC 2909
Qy 2884 GTCTGGAAGACGCTAGCTGGTGATCCCTGGATAAAGACACTTACAGCTAAATATCCCGGG 2943
Db 2910 GTGTGGAAAACACTAGCCGGCGACCCATGGATAAAAACACTGACTGCCAAGTACCCTGGG 2969
Qy 2944 GATTTCACGGCTTCATTGGACGACTGGCAGCGCGAACACGACGCCATTATGGCACGCGTT 3003
Db 2970 AATTTCACTGCCACGATAGAGGAGTGGCAAGCAGAGCATGATGCCATCATGAGGCACATC 3029
Qy 3004 CTTGATAAGCCGCAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAG 3063
Db 3030 TTGGAGAGACCGGACCCTACCGACGTCTTCCAGAATAAGGCAAACGTGTGTTGGGCCAAG 3089
Qy 3064 GCTTTAGAGCCAGTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAGACG 3123
Db 3090 GCTTTAGTGCCGGTGCTGAAGACCGCTGGCATAGACATGACCACTGAACAATGGAACACT 3149
Qy 3124 TTGCACCCATTCAAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTTTGC 3183
Db 3150 GTGGATTATTTTGAAACGGACAAAGCTCACTCAGCAGAGATAGTATTGAACCAACTATGC 3209
Qy 3184 ACCAGGTTCTTTGGAGTAGACCTGGACAGTGGGTTATTTTCCGCTCCTACCGTCGCACTT 3243
Db 3210 GTGAGGTTCTTTGGACTCGATCTGGACTCCGGTCTATTTTCTGCACCCACTGTTCCGTTA 3269
Qy 3244 ACTTACAGGGATCAGCACTGGGATAACTCGCCAGGGAAGAACATGTATGGGCTTAATAGA 3303
Db 3270 TCCATTAGGAATAATCACTGGGATAACTCCCCGTCGCCTAACATGTACGGGCTGAATAAA 3329
Qy 3304 GAGGTAGCAAAGGAGTTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACAGGC 3363
Db 3330 GAAGTGGTCCGTCAGCTCTCTCGCAGGTACCCACAACTGCCTCGGGCAGTTGCCACTGGA 3389
Qy 3364 AGGGTAGCTGATATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTGGTT 3423
Db 3390 AGAGTCTATGACATGAACACTGGTACACTGCGCAATTATGATCCGCGCATAAACCTAGTA 3449
Qy 3424 CCATTAAATCGCCGGTTGCCCCACTCGTTGATCGTTGACCACAAAGGACAGGGTACAACT 3483
Db 3450 CCTGTAAACAGAAGACTGCCTCATGCTTTAGTCCTCCACCATAATGAACACCCACAGAGT 3509
Qy 3484 GATCACAGCGGATTCCTATCTAAGATGAAGGGCAAATCTGTGTTGGTGATCGGCGATCCT 3543
Db 3510 GACTTTTCTTCATTCGTCAGCAAATTGAAGGGCAGAACTGTCCTGGTGGTCGGGGAAAAG 3569

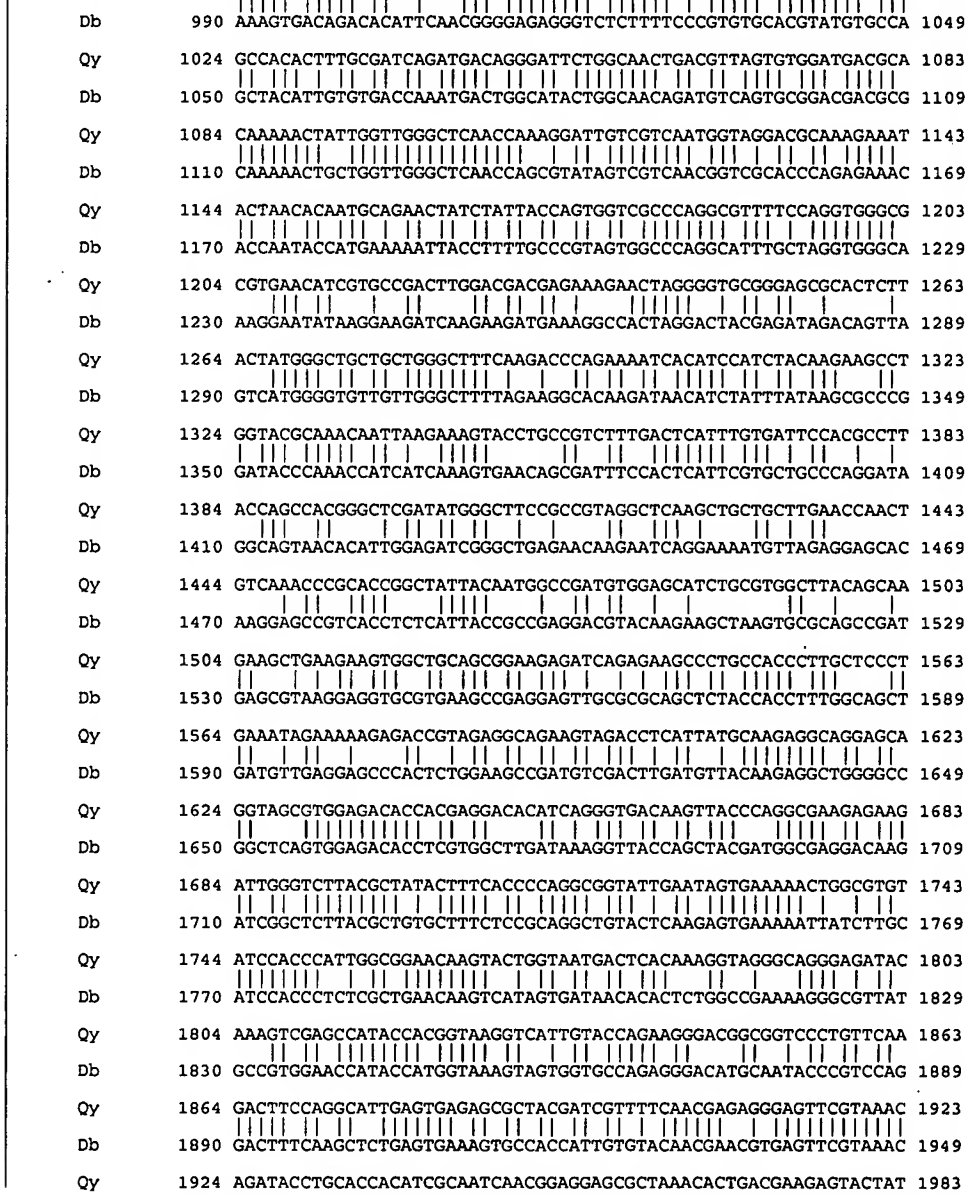
```
Qy       3544 ATCAGCATTCCAGGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATACCATCAGG 3603
Db       3570 TTGTCCGTCCCAGGCAAAATGGTTGACTGGTTGTCAGACCGGCCTGAGGCTACCTTCAGA 3629

Qy       3604 TGTGATCTCGATTTGGGAATACCTAGCCATGTCGGTAAATATGACATTATCTTTGTCAAT 3663
Db       3630 GCTCGGCTGGATTTAGGCATCCCAGGTGATGTGCCCAAATATGACATAATATTTGTTAAT 3689

Qy       3664 GTTAGGACCCCGTACAGGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCACCAC 3723
Db       3690 GTGAGGACCCCATATAAATACCATCACTATCAGCAGTGTGAAGACCATGCCATTAAGCTT 3749

Qy       3724 AGCATGCTAACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCTATA 3783
Db       3750 AGCATGTTGACCAAGAAAGCTTGTCTGCATCTGAATCCCGGCGGAACCTGTGTCAGCATA 3809

Qy       3784 GGGTATGGGCTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCACGCTCATTT 3843
Db       3810 GGTTATGGTTACGCTGACAGGGCCAGCGAAAGCATCATTGGTGCTATAGCGCGGCAGTTC 3869

Qy       3844 AGGTTTACCCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTCGTG 3903
Db       3870 AAGTTTTCCCGGGTATGCAAACCGAAATCCTCACTTGAAGAGACGGAAGTTCTGTTTGTA 3929

Qy       3904 TTCTTCGGCAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTGCTT 3963
Db       3930 TTCATTGGGTACGATCGCAAGGCCCGTACGCACAATCCTTACAAGCTTTCATCAACCTTG 3989

Qy       3964 GACAACATCTATCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGTACAGAGTG 4023
Db       3990 ACCAACATTTATACAGGTTCCAGACTCCACGAAGCCGGATGTGCACCCTCATATCATGTG 4049

Qy       4024 ATCAGAGGTGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGCAAA 4083
Db       4050 GTGCGAGGGGATATTGCCACGGCCACCGAAGGAGTGATTATAAATGCTGCTAACAGCAAA 4109

Qy       4084 GGTCAACCAGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTTGAT 4143
Db       4110 GGACAACCTGGCGGAGGGGTGTGCGGAGCGCTGTATAAGAAATTCCCGGAAAGCTTCGAT 4169

Qy       4144 AGACAGCCAATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACC---GCTCATCATA 4200
Db       4170 TTACAGCCGATCGAAGTAGGAAAAGCGCGACTGGTCAAAGGTGCAGCTAAACATATCATT 4229

Qy       4201 CATGCTGTAGGACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTAAGCTCGCA 4260
Db       4230 CATGCCGTAGGACCAAACTTCAACAAAGTTTCGGAGGTTGAAGGTGACAAACAGTTGGCA 4289

Qy       4261 GCTGCCTACATGAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTA 4320
Db       4290 GAGGCTTATGAGTCCATCGCTAAGATTGTCAACGATAACAATTACAAGTCAGTAGCGATT 4349

Qy       4321 CCGCTACTGTCAACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCAT 4380
Db       4350 CCACTGTTGTCCACCGGCATCTTTTCCGGGAACAAAGATCGACTAACCCAATCATTGAAC 4409

Qy       4381 CACCTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAA 4440
Db       4410 CATTTGCTGACAGCTTTAGACACCACTGATGCAGATGTAGCCATATACTGCAGGGACAAG 4469

Qy       4441 CAATGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATTCTGGA- 4499
Db       4470 AAATGGGAAATGACTCTCAAGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGGAGATATGC 4529

Qy       4500 --------------TGATGACAAGCCAGTAGACATTGACTTGGTCAGGGTCCACCCAAAC 4545
Db       4530 ATATCCGACGACTCTTCAGTGACAGAACCTGATGCAGAGCTGGTGAGGGTGCATCCGAAG 4589

Qy       4546 AGCTCTTTGGCAGGCAGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTG 4605
Db       4590 AGTTCTTTGGCTGGAAGGAAGGGCTACAGCACAAGCGATGGCAAAACTTTCTCATATTTG 4649

Qy       4606 GAAGGTACACGATTCCATCAGACCGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCC 4665
Db       4650 GAAGGGACCAAGTTTCACCAGGCGGCCAAGGATATAGCAGAAATTAATGCCATGTGGCCC 4709

Qy       4666 AACAAATCTGAGGCTAATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGC 4725
Db       4710 GTTGCAACGGAGGCCAATGAGCAGGTATGCATGTATATCCTCGGAGAAAGCATGAGCAGT 4769

Qy       4726 ATCCGCTCCAAATGCCCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCA 4785
Db       4770 ATTAGGTCGAAATGCCCCGTCGAAGAGTCGGAAGCCTCCTCACCACCTAGCACGCTGCCT 4829

Qy       4786 TGCCTGTGTAATTACGCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAA 4845
Db       4830 TGCTTGTGCATCCATGCCATGACTCCAGAAAGAGTACAGCGCCTAAAAGCCTCACGTCCA 4889

Qy       4846 GAACAGTTCGCCGTATGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAG 4905
Db       4890 GAACAAATTACTGTGTGCTCATCCTTTCCATTGCCGAAGTATAGAATCACTGGTGTGCAG 4949

Qy       4906 AAGCTACAGTGCAGCAAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCC 4965
Db       4950 AAGATCCAATGCTCCCAGCCTATATTGTTCTCACCGAAAGTGCCTGCGTATATTCATCCA 5009

Qy       4966 AGGAAGTA 4973
Db       5010 AGGAAGTA 5017
```

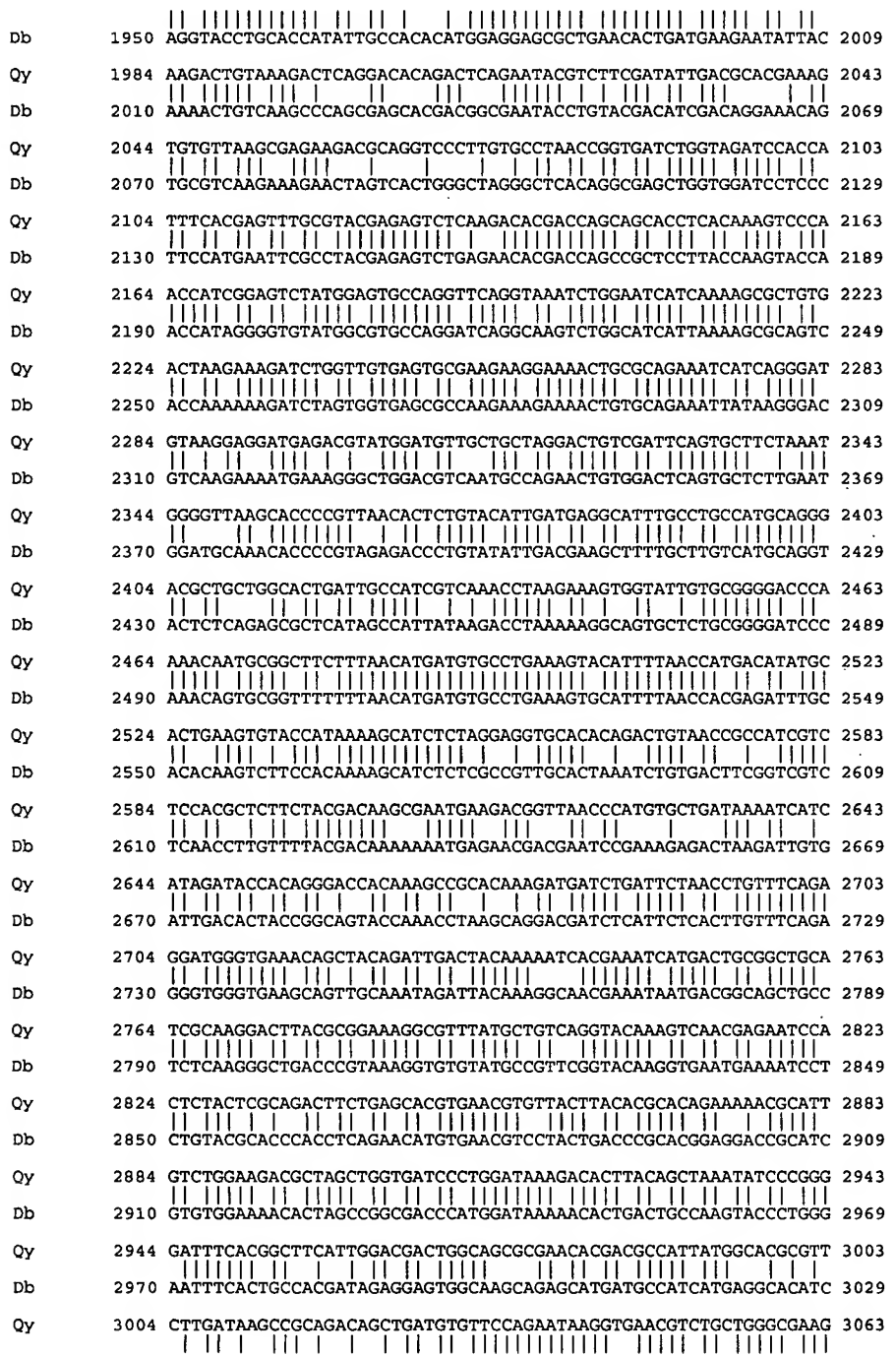
RESULT 5
AAD29141
ID AAD29141 standard; DNA; 13584 BP.
XX
AC AAD29141;
XX
DT 07-MAY-2002 (first entry)
XX
DE Venezuelan equine encephalitis virus (VEE) RNA replicon plasmid.
XX
KW Venezuelan equine encephalitis virus; VEE; vaccine; HIV infection;
KW human immunodeficiency virus; alpha-virus replicon; immune response;
KW therapy; ds.
XX
OS Chimeric - Venezuelan equine encephalitis virus.
OS Chimeric - Unidentified.
XX
PN WO200203917-A2.
XX
PD 17-JAN-2002.
XX
PF 09-JUL-2001; 2001WO-US21701.
XX
PR 07-JUL-2000; 2000US-216995P.
XX
PA (ALPH-) ALPHAVAX INC.
PA (UYNC-) UNIV NORTH CAROLINA.
XX
PI Olmsted R, Keith P, Dryga S, Caley I, Maughan M, Johnston R;
PI Davis N, Swanstrom R;
XX
DR WPI; 2002-171664/22.
XX
PT Composition useful for treating or preventing HIV infections, comprises
PT two or more isolated nucleic acids encoding env, gag or pol gene
PT product of HIV or immunogenic fragment of the gene products -
XX
PS Example 5; Page 190-195; 201pp; English.
XX
CC The invention relates to a composition comprising isolated nucleic acids
CC encoding env, gag or pol gene product of human immunodeficiency virus or
CC immunogenic fragment of the gene products. The gag gene product is
CC modified to inhibit formation of virus-like particles containing gag
CC gene product and their release from cells, and the pol gene product is
CC modified to inhibit reverse transcriptase activity. The invention also

CC relates to a method for producing an alpha-virus replicon particle used
CC in vaccines. The composition is useful for inducing an immune response
CC to human immunodeficiency virus (HIV) or for treating or preventing HIV
CC infection in a subject. The alpha-virus replicon particle is useful in a
CC vaccine. The composition is useful for administering a protein or peptide
CC to a subject. A composition comprising heparin affinity-purified alpha-
CC virus replicon particle is useful as a clinical trial material and as a
CC commercial product. The present sequence is Venezuelan equine
CC encephalitis virus (VEE) RNA replicon plasmid used in the exemplification
CC of the invention.
XX
SQ Sequence 13584 BP; 3942 A; 3114 C; 3392 G; 3136 T; 0 other;

Query Match 42.3%; Score 2113.2; DB 24; Length 13584;
Best Local Similarity 64.5%; Pred. No. 0;
Matches 3216; Conservative 0; Mismatches 1748; Indels 24; Gaps 3;





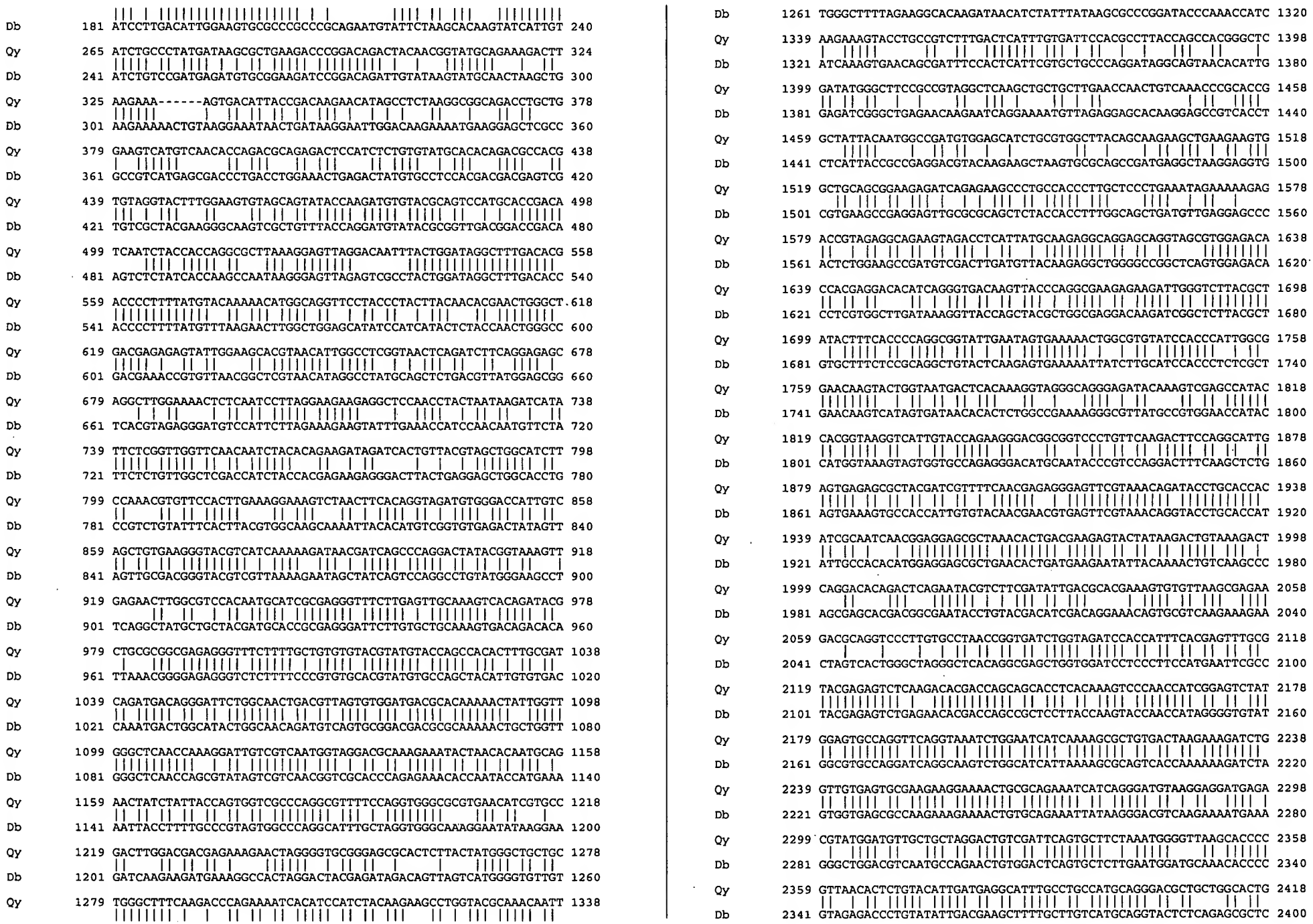

Db 1950 AGGTACCTGCACCATATTGCCACACATGGAGGAGCGCTGAACACTGATGAAGAATATTAC 2009
Qy 1984 AAGACTGTAAAGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAG 2043
Db 2010 AAAACTGTCAAGCCCAGCGAGCACGACGGCGAATACCTGTACGACATCGACAGGAAACAG 2069
Qy 2044 TGTGTTAAGCGAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCA 2103
Db 2070 TGCGTCAAGAAAGAACTAGTCACTGGGCTAGGGCTCACAGGCGAGCTGGTGGATCCTCCC 2129
Qy 2104 TTTCACGAGTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCA 2163
Db 2130 TTCCATGAATTCGCCTACGAGAGTCTGAGAACACGACCAGCCGCTCCTTACCAAGTACCA 2189
Qy 2164 ACCATCGGAGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTG 2223
Db 2190 ACCATAGGGGTGTATGGCGTGCCAGGATCAGGCAAGTCTGGCATCATTAAAAGCGCAGTC 2249
Qy 2224 ACTAAGAAAGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGAT 2283
Db 2250 ACCAAAAAAGATCTAGTGGTGAGCGCCAAGAAAGAAAACTGTGCAGAAATTATAAGGGAC 2309
Qy 2284 GTAAGGAGGATGAGACGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAAT 2343
Db 2310 GTCAAGAAAATGAAAGGGCTGGACGTCAATGCCAGAACTGTGGACTCAGTGCTCTTGAAT 2369
Qy 2344 GGGGTTAAGCACCCCGTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGG 2403
Db 2370 GGATGCAAACACCCCGTAGAGACCCTGTATATTGACGAAGCTTTTGCTTGTCATGCAGGT 2429
Qy 2404 ACGCTGCTGGCACTGATTGCCATCGTCAAACCTAAGAAAGTGGTATTGTGCGGGGACCCA 2463
Db 2430 ACTCTCAGAGCGCTCATAGCCATTATAAGACCTAAAAAGGCAGTGCTCTGCGGGGATCCC 2489
Qy 2464 AAACAATGCGGCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGC 2523
Db 2490 AAACAGTGCGGTTTTTTTAACATGATGTGCCTGAAAGTGCATTTTAACCACGAGATTTGC 2549
Qy 2524 ACTGAAGTGTACCATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATCGTC 2583
Db 2550 ACACAAGTCTTCCACAAAAGCATCTCTCGCCGTTGCACTAAATCTGTGACTTCGGTCGTC 2609
Qy 2584 TCCACGCTCTTCTACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATCATC 2643
Db 2610 TCAACCTTGTTTTACGACAAAAAAATGAGAACGACGAATCCGAAAGAGACTAAGATTGTG 2669
Qy 2644 ATAGATACCACAGGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGA 2703
Db 2670 ATTGACACTACCGGCAGTACCAAACCTAAGCAGGACGATCTCATTCTCACTTGTTTCAGA 2729
Qy 2704 GGATGGGTGAAACAGCTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCTGCA 2763
Db 2730 GGGTGGGTGAAGCAGTTGCAAATAGATTACAAAGGCAACGAAATAATGACGGCAGCTGCC 2789
Qy 2764 TCGCAAGGACTTACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCA 2823
Db 2790 TCTCAAGGGCTGACCCGTAAAGGTGTGTATGCCGTTCGGTACAAGGTGAATGAAAATCCT 2849
Qy 2824 CTCTACTCGCAGACTTCTGAGCACGTGAACGTGTTACTTACACGCACAGAAAAACGCATT 2883
Db 2850 CTGTACGCACCCACCTCAGAACATGTGAACGTCCTACTGACCCGCACGGAGGACCGCATC 2909
Qy 2884 GTCTGGAAGACGCTAGCTGGTGATCCCTGGATAAAGACACTTACAGCTAAATATCCCGGG 2943
Db 2910 GTGTGGAAAACACTAGCCGGCGACCCATGGATAAAAACACTGACTGCCAAGTACCCTGGG 2969
Qy 2944 GATTTCACGGCTTCATTGGACGACTGGCAGCGCGAACACGACGCCATTATGGCACGCGTT 3003
Db 2970 AATTTCACTGCCACGATAGAGGAGTGGCAAGCAGAGCATGATGCCATCATGAGGCACATC 3029
Qy 3004 CTTGATAAGCCGCAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAG 3063
Db 3030 TTGGAGAGACCGGACCCTACCGACGTCTTCCAGAATAAGGCAAACGTGTGTTGGGCCAAG 3089
Qy 3064 GCTTTAGAGCCAGTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAGACG 3123
Db 3090 GCTTTAGTGCCGGTGCTGAAGACCGCTGGCATAGACATGACCACTGAACAATGGAACACT 3149
Qy 3124 TTGCACCCATTCAAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTTTGC 3183
Db 3150 GTGGATTATTTTGAAACGGACAAAGCTCACTCAGCAGAGATAGTATTGAACCAACTATGC 3209
Qy 3184 ACCAGGTTCTTTGGAGTAGACCTGGACAGTGGGTTATTTTCCGCTCCTACCGTCGCACTT 3243
Db 3210 GTGAGGTTCTTTGGACTCGATCTGGACTCCGGTCTATTTTCTGCACCCACTGTTCCGTTA 3269
Qy 3244 ACTTACAGGGATCAGCACTGGGATAACTCGCCAGGGAAGAACATGTATGGGCTTAATAGA 3303
Db 3270 TCCATTAGGAATAATCACTGGGATAACTCCCCGTCGCCTAACATGTACGGGCTGAATAAA 3329
Qy 3304 GAGGTAGCAAAGGAGTTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACAGGC 3363
Db 3330 GAAGTGGTCCGTCAGCTCTCTCGCAGGTACCCACAACTGCCTCGGGCAGTTGCCACTGGA 3389
Qy 3364 AGGGTAGCTGATATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTGGTT 3423
Db 3390 AGAGTCTATGACATGAACACTGGTACACTGCGCAATTATGATCCGCGCATAAACCTAGTA 3449
Qy 3424 CCATTAAATCGCCGGTTGCCCCACTCGTTGATCGTTGACCACAAAGGACAGGGTACAACT 3483
Db 3450 CCTGTAAACAGAAGACTGCCTCATGCTTTAGTCCTCCACCATAATGAACACCCACAGAGT 3509
Qy 3484 GATCACAGCGGATTCCTATCTAAGATGAAGGGCAAATCTGTGTTGGTGATCGGCGATCCT 3543
Db 3510 GACTTTTCTTCATTCGTCAGCAAATTGAAGGGCAGAACTGTCCTGGTGGTCGGGGAAAAG 3569
Qy 3544 ATCAGCATTCCAGGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATACCATCAGG 3603
Db 3570 TTGTCCGTCCCAGGCAAAATGGTTGACTGGTTGTCAGACCGGCCTGAGGCTACCTTCAGA 3629
Qy 3604 TGTGATCTCGATTTGGGAATACCTAGCCATGTCGGTAAATATGACATTATCTTTGTCAAT 3663
Db 3630 GCTCGGCTGGATTTAGGCATCCCAGGTGATGTGCCCAAATATGACATAATATTTGTTAAT 3689
Qy 3664 GTTAGGACCCCGTACAGGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCACCAC 3723
Db 3690 GTGAGGACCCCATATAAATACCATCACTATCAGCAGTGTGAAGACCATGCCATTAAGCTT 3749
Qy 3724 AGCATGCTAACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCTATA 3783
Db 3750 AGCATGTTGACCAAGAAAGCTTGTCTGCATCTGAATCCCGGCGGAACCTGTGTCAGCATA 3809
Qy 3784 GGGTATGGGCTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCACGCTCATTT 3843
Db 3810 GGTTATGGTTACGCTGACAGGGCCAGCGAAAGCATCATTGGTGCTATAGCGCGGCAGTTC 3869
Qy 3844 AGGTTTACCCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTCGTG 3903
Db 3870 AAGTTTTCCCGGGTATGCAAACCGAAATCCTCACTTGAAGAGACGGAAGTTCTGTTTGTA 3929
Qy 3904 TTCTTCGGCAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTGCTT 3963
Db 3930 TTCATTGGGTACGATCGCAAGGCCCGTACGCACAATCCTTACAAGCTTTCATCAACCTTG 3989
Qy 3964 GACAACATCTATCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGTACAGAGTG 4023
Db 3990 ACCAACATTTATACAGGTTCCAGACTCCACGAAGCCGGATGTGCACCCTCATATCATGTG 4049
Qy 4024 ATCAGAGGTGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGCAAA 4083
Db 4050 GTGCGAGGGGATATTGCCACGGCCACCGAAGGAGTGATTATAAATGCTGCTAACAGCAAA 4109
Qy 4084 GGTCAACCAGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTTGAT 4143
Db 4110 GGACAACCTGGCGGAGGGGTGTGCGGAGCGCTGTATAAGAAATTCCCGGAAAGCTTCGAT 4169

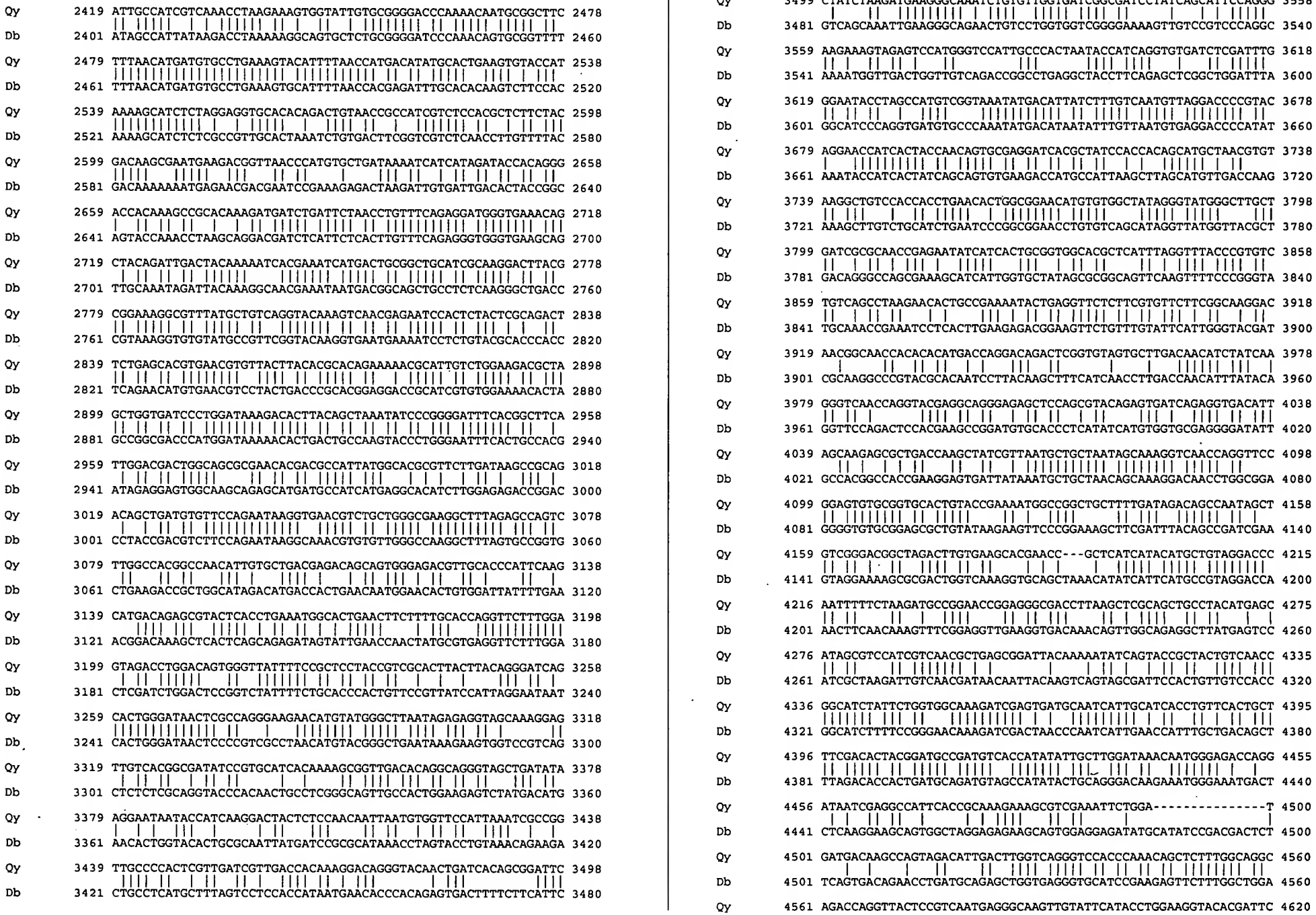
```
Qy       4144 AGACAGCCAATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACC---GCTCATCATA 4200
Db       4170 TTACAGCCGATCGAAGTAGGAAAAGCGCGACTGGTCAAAGGTGCAGCTAAACATATCATT 4229

Qy       4201 CATGCTGTAGGACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTAAGCTCGCA 4260
Db       4230 CATGCCGTAGGACCAAACTTCAACAAAGTTTCGGAGGTTGAAGGTGACAAACAGTTGGCA 4289

Qy       4261 GCTGCCTACATGAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTA 4320
Db       4290 GAGGCTTATGAGTCCATCGCTAAGATTGTCAACGATAACAATTACAAGTCAGTAGCGATT 4349

Qy       4321 CCGCTACTGTCAACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCAT 4380
Db       4350 CCACTGTTGTCCACCGGCATCTTTTCCGGGAACAAAGATCGACTAACCCAATCATTGAAC 4409

Qy       4381 CACCTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAA 4440
Db       4410 CATTTGCTGACAGCTTTAGACACCACTGATGCAGATGTAGCCATATACTGCAGGGACAAG 4469

Qy       4441 CAATGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATTCTGGA- 4499
Db       4470 AAATGGGAAATGACTCTCAAGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGGAGATATGC 4529

Qy       4500 --------------TGATGACAAGCCAGTAGACATTGACTTGGTCAGGGTCCACCCAAAC 4545
Db       4530 ATATCCGACGACTCTTCAGTGACAGAACCTGATGCAGAGCTGGTGAGGGTGCATCCGAAG 4589

Qy       4546 AGCTCTTTGGCAGGCAGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTG 4605
Db       4590 AGTTCTTTGGCTGGAAGGAAGGGCTACAGCACAAGCGATGGCAAAACTTTCTCATATTTG 4649

Qy       4606 GAAGGTACACGATTCCATCAGACCGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCC 4665
Db       4650 GAAGGGACCAAGTTTCACCAGGCGGCCAAGGATATAGCAGAAATTAATGCCATGTGGCCC 4709

Qy       4666 AACAAATCTGAGGCTAATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGC 4725
Db       4710 GTTGCAACGGAGGCCAATGAGCAGGTATGCATGTATATCCTCGGAGAAAGCATGAGCAGT 4769

Qy       4726 ATCCGCTCCAAATGCCCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCA 4785
Db       4770 ATTAGGTCGAAATGCCCCGTCGAAGAGTCGGAAGCCTCCTCACCACCTAGCACGCTGCCT 4829

Qy       4786 TGCCTGTGTAATTACGCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAA 4845
Db       4830 TGCTTGTGCATCCATGCCATGACTCCAGAAAGAGTACAGCGCCTAAAAGCCTCACGTCCA 4889

Qy       4846 GAACAGTTCGCCGTATGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAG 4905
Db       4890 GAACAAATTACTGTGTGCTCATCCTTTCCATTGCCGAAGTATAGAATCACTGGTGTGCAG 4949

Qy       4906 AAGCTACAGTGCAGCAAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCC 4965
Db       4950 AAGATCCAATGCTCCCAGCCTATATTGTTCTCACCGAAAGTGCCTGCGTATATTCATCCA 5009

Qy       4966 AGGAAGTA 4973
Db       5010 AGGAAGTA 5017
```

RESULT 6
AAD29132
ID AAD29132 standard; DNA; 7479 BP.
XX
AC AAD29132;
XX
DT 07-MAY-2002 (first entry)
XX
DE Venezuelan equine encephalitis virus (VEE) non-structural gene.
XX
KW Venezuelan equine encephalitis virus; VEE; vaccine; HIV infection;
KW human immunodeficiency virus; alpha-virus replicon; immune response;
KW therapy; non-structural gene; ds.
XX
OS Venezuelan equine encephalitis virus.
XX
FH Key Location/Qualifiers
FT CDS 1..7479
FT /*tag= a
FT /product= "VEE non-structural protein"
FT /transl_except= (pos:5635..5643, aa:Gln-Arg)
FT /note= "Insertion of 3 bases alters the reading frame;
FT CDS does not include stop codon"
FT /partial
XX
PN WO200203917-A2.
XX
PD 17-JAN-2002.
XX
PF 09-JUL-2001; 2001WO-US21701.
XX
PR 07-JUL-2000; 2000US-216995P.
XX
PA (ALPH-) ALPHAVAX INC.
PA (UYNC-) UNIV NORTH CAROLINA.
XX
PI Olmsted R, Keith P, Dryga S, Caley I, Maughan M, Johnston R;
PI Davis N, Swanstrom R;
XX
DR WPI; 2002-171664/22.
DR P-PSDB; AAE18302.
XX
PT Composition useful for treating or preventing HIV infections, comprises
PT two or more isolated nucleic acids encoding env, gag or pol gene
PT product of HIV or immunogenic fragment of the gene products -
XX
PS Example 2; Page 145-156; 201pp; English.
XX
CC The invention relates to a composition comprising isolated nucleic acids
CC encoding env, gag or pol gene product of human immunodeficiency virus or
CC immunogenic fragment of the gene products. The gag gene product is
CC modified to inhibit formation of virus-like particles containing gag
CC gene product and their release from cells, and the pol gene product is
CC modified to inhibit reverse transcriptase activity. The invention also
CC relates to a method for producing an alpha-virus replicon particle used
CC in vaccines. The composition is useful for inducing an immune response
CC to human immunodeficiency virus (HIV) or for treating or preventing HIV
CC infection in a subject. The alpha-virus replicon particle is useful in a
CC vaccine. The composition is useful for administering a protein or peptide
CC to a subject. A composition comprising heparin affinity-purified alpha-
CC virus replicon particle is useful as a clinical trial material and as a
CC commercial product. The present sequence is a non-structural gene from
CC Venezuelan equine encephalitis virus (VEE) used in the exemplification
CC of the invention.
XX
SQ Sequence 7479 BP; 2141 A; 1774 C; 1900 G; 1664 T; 0 other;

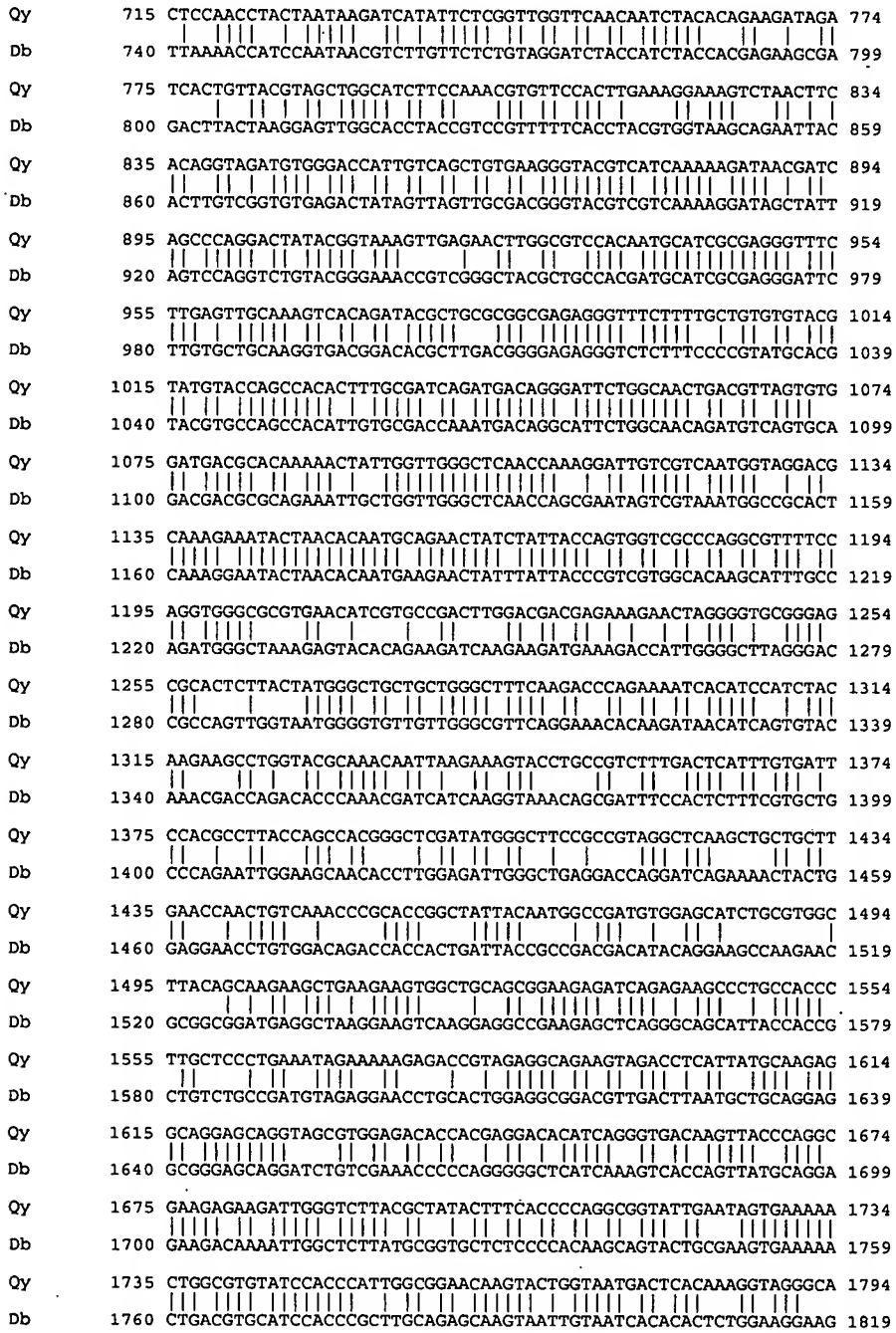
Query Match 42.2%; Score 2111; DB 24; Length 7479;
Best Local Similarity 64.5%; Pred. No. 0;
Matches 3209; Conservative 0; Mismatches 1740; Indels 24; Gaps 3;

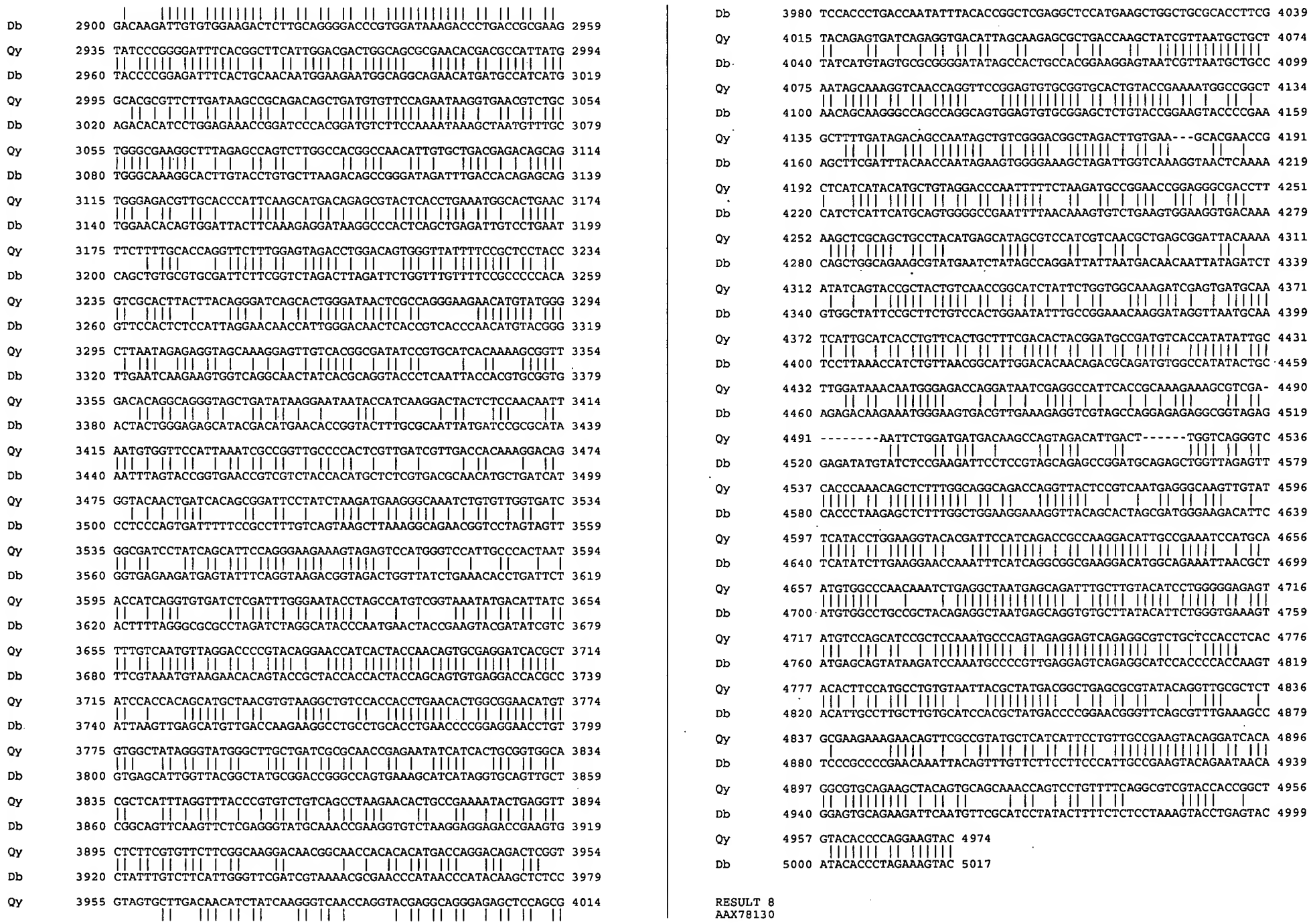
```
Qy         25 ATGGAAAGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGTCGTTACAGCGG 84
Db          1 ATGGAGAAAGTTCACGTTGACATCGAGGAAGACAGCCCATTCCTCAGAGCTTTGCAGCGG 60

Qy         85 ACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACCATGCCAATGCC 144
Db         61 AGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCACTGATAATGACCATGCTAATGCC 120

Qy        145 AGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACCGGGACCAAGTT 204
Db        121 AGAGCGTTTTCGCATCTGGCTTCAAAACTGATCGAAACGGAGGTGGACCCATCCGACACG 180

Qy        205 ATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACCGCTATCATTGT 264
```





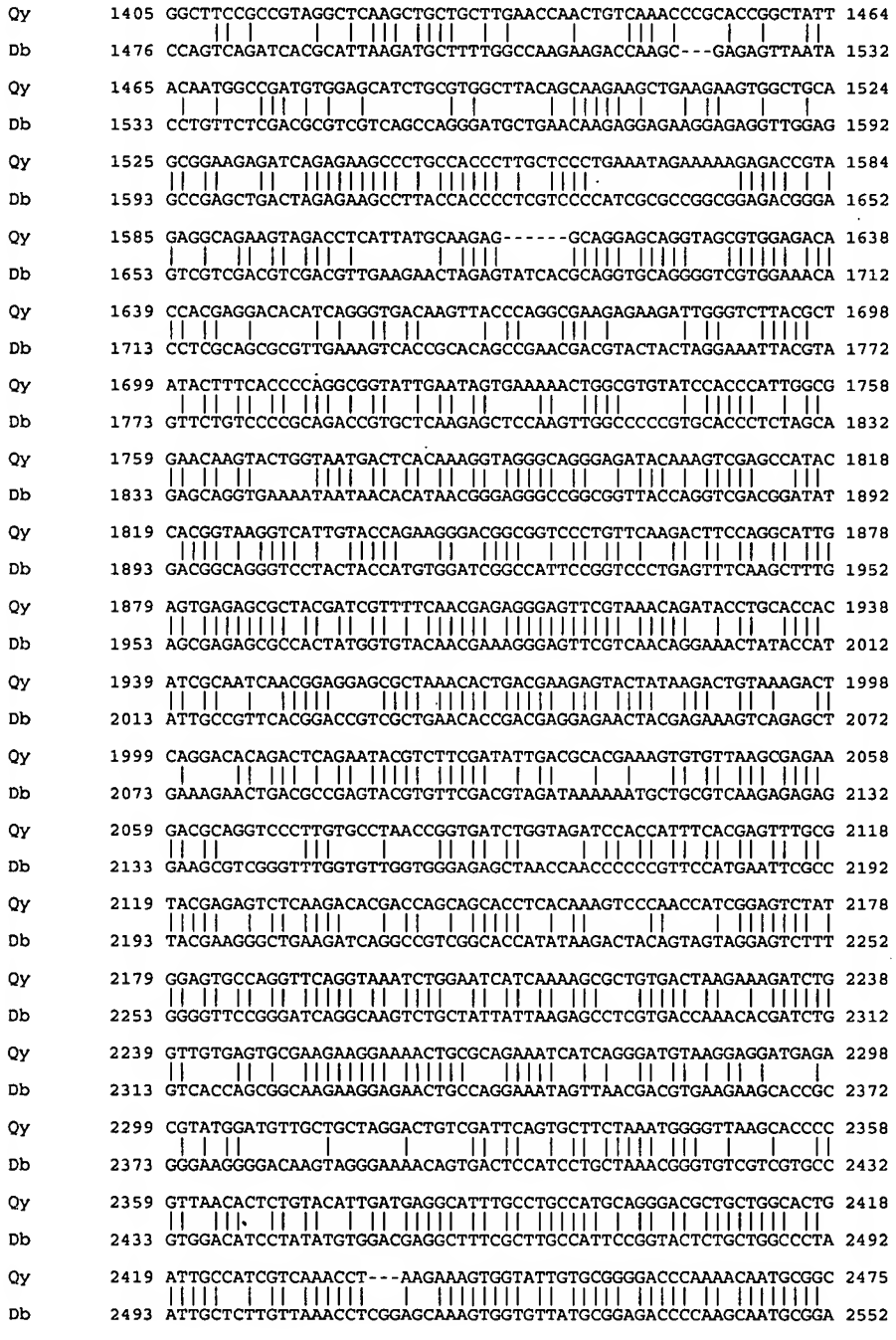
```
Db       4561 AGGAAGGGCTACAGCACAAGCGATGGCAAAACTTTCTCATATTTGGAAGGGACCAAGTTT 4620

Qy       4621 CATCAGACCGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCCAACAAATCTGAGGCT 4680

Db       4621 CACCAGGCGGCCAAGGATATAGCAGAAATTAATGCCATGTGGCCCGTTGCAACGGAGGCC 4680

Qy       4681 AATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGC 4740

Db       4681 AATGAGCAGGTATGCATGTATATCCTCGGAGAAAGCATGAGCAGTATTAGGTCGAAATGC 4740

Qy       4741 CCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTAC 4800

Db       4741 CCCGTCGAAGAGTCGGAAGCCTCCACACCACCTAGCACGCTGCCTTGCTTGTGCATCCAT 4800

Qy       4801 GCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTA 4860

Db       4801 GCCATGACTCCAGAAAGAGTACAGCGCCTAAAAGCCTCACGTCCAGAACAAATTACTGTG 4860

Qy       4861 TGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGC 4920

Db       4861 TGCTCATCCTTTCCATTGCCGAAGTATAGAATCACTGGTGTGCAGAAGATCCAATGCTCC 4920

Qy       4921 AAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTA 4973

Db       4921 CAGCCTATATTGTTCTCACCGAAAGTGCCTGCGTATATTCATCCAAGGAAGTA 4973


RESULT 7
AAV74108
ID   AAV74108 standard; cDNA; 11464 BP.
XX
AC   AAV74108;
XX
DT   12-APR-1999  (first entry)
XX
DE   Venezuelan equine encephalitis virus IE variant cDNA.
XX
KW   VEE IE virus; vaccine; ds.
XX
OS   Venezuelan equine encephalitis virus.
XX
PN   WO9853077-A1.
XX
PD   26-NOV-1998.
XX
PF   20-MAY-1998;   98WO-US10645.
XX
PR   16-DEC-1997;   97US-0991840.
PR   20-MAY-1997;   97US-0047162.
PR   24-JUL-1997;   97US-0053652.
XX
PA   (REED-) REED ARMY INST RES WALTER.
XX
PI   Crise BJ,  Oberste MS,  Parker MD,  Schmura SM,  Smith JF;
XX
DR   WPI; 1999-045316/04.
XX
PT   New DNA encoding infectious Western or Venezuelan equine
PT   encephalitis virus genome - useful for the production of live or
PT   attenuated vaccines for human or veterinary medicine
XX
PS   Claim 24; Page 68-79; 112pp; English.
XX
CC   This cDNA sequence codes for an infectious Venezuelan equine
CC   encephalitis virus IE variant (VEE IE) RNA transcript.  Using
CC   primers (see AAV74122-29) specific to genomic RNA of VEE IE, RT-PCR
CC   was used to generate cDNA fragments which were subsequently cloned
CC   and used to assemble the full-length cDNA.  The full-length
CC   infectious clone is useful in the production of virulent VEE IE
CC   virus, and for introducing and testing attenuating mutations.  Also
CC   new are: infectious or attenuated VEE IE RNA transcripts and viral
CC   particles; cDNA (see AAV74107) encoding an infectious Western equine
CC   encephalitis (WEE) viral genome; infectious or attenuated WEE RNA
CC   transcripts and viral particles; an attenuated chimeric virus
CC   containing non-structural sequences from a first alpha-virus (aV)
CC   and structural sequences from a second aV, resulting in attenuation
CC   of the second aV; a method for expressing a protein (especially an
CC   antigen to protect against a pathogen) by cloning its gene into an
CC   attenuated WEE or VEE IE replicon so that transcription of the
CC   replicon produces RNA able to infect the cells in which protein is
CC   to be produced; methods for diagnosing WEE and VEE infections; and
CC   polypeptides encoded by VEE variant IIIA (see AAV74109).  Attenuated
CC   WEE and VEE are used in live or inactivated vaccines, for use in
CC   human or veterinary medicine.  Chimeric viruses are also useful as
CC   vaccines, directed against the second aV which is particularly
CC   Eastern equine encephalitis (EEE) virus or some variant of VEE.
CC   WEE and VEE nucleic acids are used as primers and probes to
CC   diagnose virus infections and to define natural variants, also for
CC   production of protein antigens which can be used as diagnostic
CC   reagents, to generate antibodies, and in vaccines.  The attenuated
CC   viruses are highly immunogenic and provide long-lasting protection.
XX
SQ   Sequence 11464 BP; 3259 A; 2734 C; 2900 G; 2571 T; 0 other;

  Query Match             41.5%;  Score 2074.8;  DB 20;  Length 11464;
  Best Local Similarity   64.1%;  Pred. No. 0;
  Matches 3202;  Conservative    0;  Mismatches 1772;  Indels   24;  Gaps    4;

Qy          1 ACCCTACAAACTAATCGATCCAATATGGAAAGAATTCACGTTGACTTAGATGCTGACAGC 60

Db         20 AGCCCAAACCAATAACTACCCAAAATGGAGAAAGTTCACGTTGACATCGAGGAAGATAGT 79

Qy         61 CCGTATGTCAAGTCGTTACAGCGGACGTTTCCACAATTTGAGATCGAAGCAAGGCAGGTC 120

Db         80 CCCTTCCTCAGAGCATTACAACGGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTC 139

Qy        121 ACTGACAATGACCATGCCAATGCCAGAGCGTTTTCGCATGTGGCAACAAAGCTCATTGAG 180

Db        140 ACAGATAATGACCATGCTAACGCCAGAGCGTTTTCGCATTTGGCATCGAAATTGATCGAG 199

Qy        181 AGCGAAGTCGACCGGGACCAAGTTATCTTGGACATTGGAAGTGCGCCCGTCAGACATGCA 240

Db        200 ACGGAGGTGGAACCATCCGATACGATCCTAGACATTGGAAGTGCGCCTGCCCGCAGAATG 259

Qy        241 CATTCCAATCACCGCTATCATTGTATCTGCCCTATGATAAGCGCTGAAGACCCGGACAGA 300

Db        260 TATTCCAAGCATAAGTACCATTGCATCTGTCCGATGAAATGTGCAGAAGATCCGGACAGA 319

Qy        301 CTACAACGGTATGCAGAAAGACTTAAGAAAAGT------GACATTACCGACAAGAACATA 354

Db        320 CTGTTTAAGTATGCAGCCAAGCTGAAGAAGAACTGTAAAGAGATTACAGATAAGGAACTG 379

Qy        355 GCCTCTAAGGCGGCAGACCTGCTGGAAGTCATGTCAACACCAGACGCAGAGACTCCATCT 414

Db        380 GACAAGAAGATGAAGGAGCTTGCGGAAGTCATGAGCGACCCTGATCTCGAAACTGAAACG 439

Qy        415 CTGTGTATGCACACAGACGCCACGTGTAGGTACTTTGGAAGTGTAGCAGTATACCAAGAT 474

Db        440 ATTTGCCTTCACGACGATGAAACCTGTCGATTTGAGGGTCAAGTCGCAGTGTATCAGGAT 499

Qy        475 GTGTACGCAGTCCATGCACCGACATCAATCTACCACCAGGCGCTTAAAGGAGTTAGGACA 534

Db        500 GTGTACGCGGTTGACGGACCGACGAGCCTTTACCATCAGGCCAACAAAGGGGTCAGAGTC 559

Qy        535 ATTTACTGGATAGGCTTTGACACGACCCCTTTTATGTACAAAAACATGGCAGGTTCCTAC 594

Db        560 GCCTATTGGATAGGATTCGACACTACCCCTTTTATGTTTAAGAACCTGGCTGGAGCATAT 619

Qy        595 CCTACTTACAACACGAACTGGGCTGACGAGAGAGTATTGGAAGCACGTAACATTGGCCTC 654

Db        620 CCCTCCTATTCGACCAACTGGGCCGACGAGACCGTGTTAACGGCTCGTAATATAGGCTTG 679

Qy        655 GGTAACTCAGATCTTCAGGAGAGCAGGCTTGGAAAACTCTCAATCCTTAGGAAGAAGAGG 714

Db        680 TGCAGCTCCGATGTGATGGAGCGGTCACGTAGAGGGATGTCCATCCTCAGAAAGAAATTT 739
```

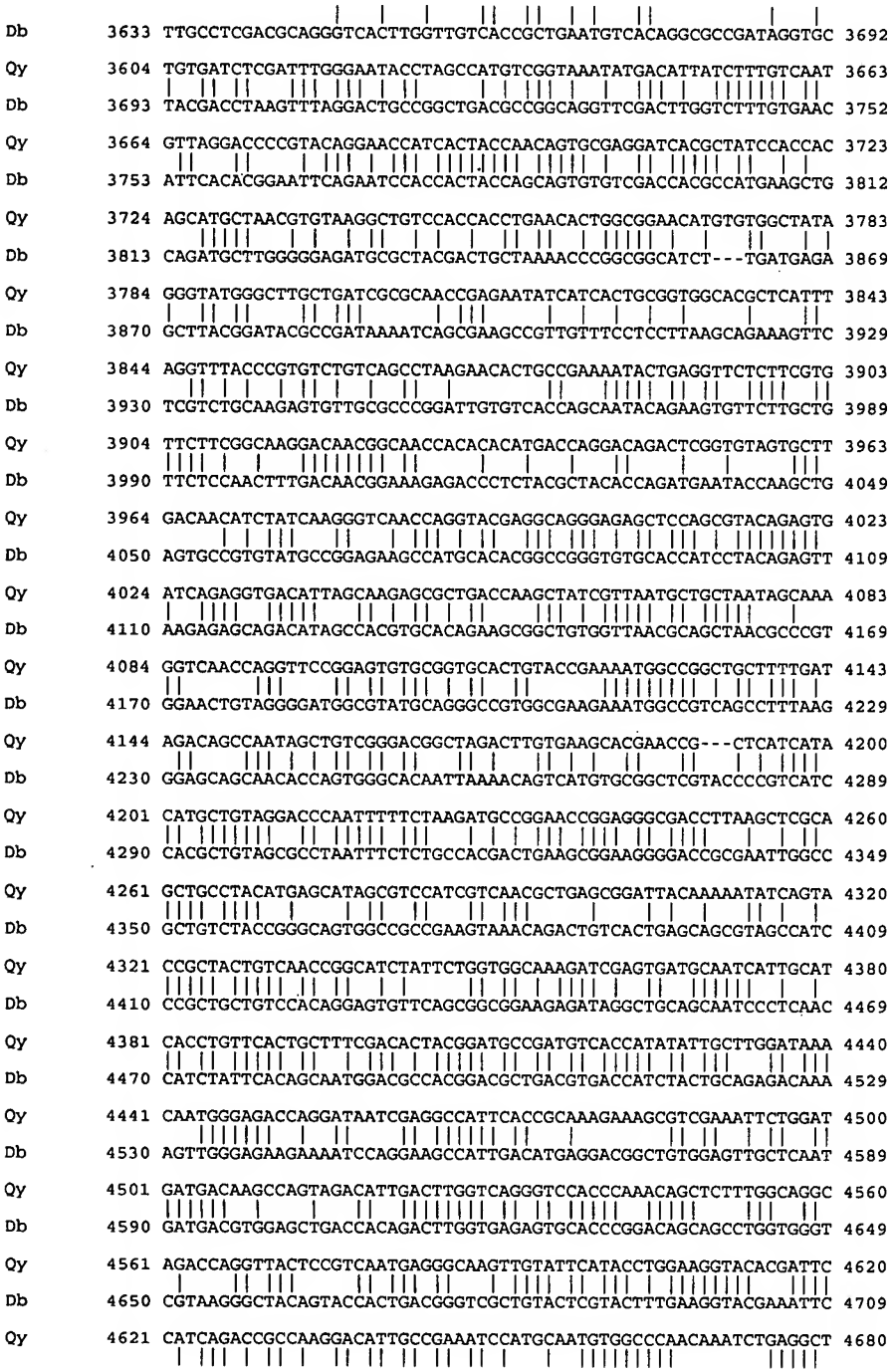

Qy 715 CTCCAACCTACTAATAAGATCATATTCTCGGTTGGTTCAACAATCTACACAGAAGATAGA 774
Db 740 TTAAAACCATCCAATAACGTCTTGTTCTCTGTAGGATCTACCATCTACCACGAGAAGCGA 799
Qy 775 TCACTGTTACGTAGCTGGCATCTTCCAAACGTGTTCCACTTGAAAGGAAAGTCTAACTTC 834
Db 800 GACTTACTAAGGAGTTGGCACCTACCGTCCGTTTTTCACCTACGTGGTAAGCAGAATTAC 859
Qy 835 ACAGGTAGATGTGGGACCATTGTCAGCTGTGAAGGGTACGTCATCAAAAAGATAACGATC 894
Db 860 ACTTGTCGGTGTGAGACTATAGTTAGTTGCGACGGGTACGTCGTCAAAAGGATAGCTATT 919
Qy 895 AGCCCAGGACTATACGGTAAAGTTGAGAACTTGGCGTCCACAATGCATCGCGAGGGTTTC 954
Db 920 AGTCCAGGTCTGTACGGGAAACCGTCGGGCTACGCTGCCACGATGCATCGCGAGGGATTC 979
Qy 955 TTGAGTTGCAAAGTCACAGATACGCTGCGCGGCGAGAGGGTTTCTTTTGCTGTGTGTACG 1014
Db 980 TTGTGCTGCAAGGTGACGGACACGCTTGACGGGGAGAGGGTCTCTTTCCCCGTATGCACG 1039
Qy 1015 TATGTACCAGCCACACTTTGCGATCAGATGACAGGGATTCTGGCAACTGACGTTAGTGTG 1074
Db 1040 TACGTGCCAGCCACATTGTGCGACCAAATGACAGGCATTCTGGCAACAGATGTCAGTGCA 1099
Qy 1075 GATGACGCACAAAAACTATTGGTTGGGCTCAACCAAAGGATTGTCGTCAATGGTAGGACG 1134
Db 1100 GACGACGCGCAGAAATTGCTGGTTGGGCTCAACCAGCGAATAGTCGTAAATGGCCGCACT 1159
Qy 1135 CAAAGAAATACTAACACAATGCAGAACTATCTATTACCAGTGGTCGCCCAGGCGTTTTCC 1194
Db 1160 CAAAGGAATACTAACACAATGAAGAACTATTTATTACCCGTCGTGGCACAAGCATTTGCC 1219
Qy 1195 AGGTGGGCGCGTGAACATCGTGCCGACTTGGACGACGAGAAAGAACTAGGGGTGCGGGAG 1254
Db 1220 AGATGGGCTAAAGAGTACACAGAAGATCAAGAAGATGAAAGACCATTGGGGCTTAGGGAC 1279
Qy 1255 CGCACTCTTACTATGGGCTGCTGCTGGGCTTTCAAGACCCAGAAAATCACATCCATCTAC 1314
Db 1280 CGCCAGTTGGTAATGGGGTGTTGTTGGGCGTTCAGGAAACACAAGATAACATCAGTGTAC 1339
Qy 1315 AAGAAGCCTGGTACGCAAACAATTAAGAAAGTACCTGCCGTCTTTGACTCATTTGTGATT 1374
Db 1340 AAACGACCAGACACCCAAACGATCATCAAGGTAAACAGCGATTTCCACTCTTTCGTGCTG 1399
Qy 1375 CCACGCCTTACCAGCCACGGGCTCGATATGGGCTTCCGCCGTAGGCTCAAGCTGCTGCTT 1434
Db 1400 CCCAGAATTGGAAGCAACACCTTGGAGATTGGGCTGAGGACCAGGATCAGAAAACTACTG 1459
Qy 1435 GAACCAACTGTCAAACCCGCACCGGCTATTACAATGGCCGATGTGGAGCATCTGCGTGGC 1494
Db 1460 GAGGAACCTGTGGACAGACCACCACTGATTACCGCCGACGACATACAGGAAGCCAAGAAC 1519
Qy 1495 TTACAGCAAGAAGCTGAAGAAGTGGCTGCAGCGGAAGAGATCAGAGAAGCCCTGCCACCC 1554
Db 1520 GCGGCGGATGAGGCTAAGGAAGTCAAGGAGGCCGAAGAGCTCAGGGCAGCATTACCACCG 1579
Qy 1555 TTGCTCCCTGAAATAGAAAAAGAGACCGTAGAGGCAGAAGTAGACCTCATTATGCAAGAG 1614
Db 1580 CTGTCTGCCGATGTAGAGGAACCTGCACTGGAGGCGGACGTTGACTTAATGCTGCAGGAG 1639
Qy 1615 GCAGGAGCAGGTAGCGTGGAGACACCACGAGGACACATCAGGGTGACAAGTTACCCAGGC 1674
Db 1640 GCGGGAGCAGGATCTGTCGAAACCCCCAGGGGGCTCATCAAAGTCACCAGTTATGCAGGA 1699
Qy 1675 GAAGAGAAGATTGGGTCTTACGCTATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAA 1734
Db 1700 GAAGACAAAATTGGCTCTTATGCGGTGCTCTCCCCACAAGCAGTACTGCGAAGTGAAAAA 1759
Qy 1735 CTGGCGTGTATCCACCCATTGGCGGAACAAGTACTGGTAATGACTCACAAAGGTAGGGCA 1794
Db 1760 CTGACGTGCATCCACCCGCTTGCAGAGCAAGTAATTGTAATCACACACTCTGGAAGGAAG 1819
Qy 1795 GGGAGATACAAAGTCGAGCCATACCACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTC 1854
Db 1820 GGTAGATACGCAGTTGAGCCTTACCACGGAAAAGTGGTAGTACCCGAAGGGCAAGCTATA 1879
Qy 1855 CCTGTTCAAGACTTCCAGGCATTGAGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAG 1914
Db 1880 CCTGTCCAAGACTTCCAAGCACTCAGTGAGAGCGCCACAATTGTGTACAACGAACGTGAG 1939
Qy 1915 TTCGTAAACAGATACCTGCACCACATCGCAATCAACGGAGGAGCGCTAAACACTGACGAA 1974
Db 1940 TTCGTGAATAGGTATTTACACCACATTGCCACGCACGGTGGAGCTCTGAACACAGACGAG 1999
Qy 1975 GAGTACTATAAGACTGTAAAGACTCAGGACACAGACTCAGAATACGTCTTCGATATTGAC 2034
Db 2000 GAGTATTACAGAGTGGTGAAACCTAGTGAACATGAAGGTGAGTACTTATATGACATTGAC 2059
Qy 2035 GCACGAAAGTGTGTTAAGCGAGAAGACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTA 2094
Db 2060 AAGAAACAATGTGTGAAGAAAGAACTGGTTACGGGACTAGGTTTGACAGGAGAACTTGTC 2119
Qy 2095 GATCCACCATTTCACGAGTTTGCGTACGAGAGTCTCAAGACACGACCAGCAGCACCTCAC 2154
Db 2120 GATCCCCCCTTCCATGAATTTGCATATGAAAGCTTGAGGACACGTCCGCCCGCACCTTAT 2179
Qy 2155 AAAGTCCCAACCATCGGAGTCTATGGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAA 2214
Db 2180 CAAGTACCAACTATAGGTGTGTATGGCGTTCCCGGGTCCGGGAAGTCTGGAATTATTAAG 2239
Qy 2215 AGCGCTGTGACTAAGAAAGATCTGGTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATC 2274
Db 2240 AGCGCAGTCACAAAGAAGGACTTGGTAGTGAGTGCTAAGAAGGAGAACTGCGCCGAGATA 2299
Qy 2275 ATCAGGGATGTAAGGAGGATGAGACGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTG 2334
Db 2300 ATAAGGGACGTCAAGAAGATGAAAGGGCTAGACGTCAATGCCCGGACGGTGGACTCAGTG 2359
Qy 2335 CTTCTAAATGGGGTTAAGCACCCCGTTAACACTCTGTACATTGATGAGGCATTTGCCTGC 2394
Db 2360 CTGTTAAATGGATGCAAGCACCCGGTCGAGACTTTGTACATTGATGAAGCCTTTGCATGC 2419
Qy 2395 CATGCAGGGACGCTGCTGGCACTGATTGCCATCGTCAAACCTAAGAAAGTGGTATTGTGC 2454
Db 2420 CACGCCGGCACTCTCAGGGCCTTAATAGCCATTATACGCCCAAAGAAAGCAGTGCTATGT 2479
Qy 2455 GGGGACCCAAAACAATGCGGCTTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCAT 2514
Db 2480 GGTGACCCAAAGCAATGTGGCTTCTTCAACATGATGTGCCTGAAAGTGCACTTCAACCAT 2539
Qy 2515 GACATATGCACTGAAGTGTACCATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACC 2574
Db 2540 GAAATATGCACTCAAGTTTTCCACAAAAGCATATCTCGCAGGTGTACCAAGTCAGTGACG 2599
Qy 2575 GCCATCGTCTCCACGCTCTTCTACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGAT 2634
Db 2600 TCGGTAGTGTCCACACTGTTTTATGACAAAAGAATGAGAACTACTAACCCAAGGGATTCC 2659
Qy 2635 AAAATCATCATAGATACCACAGGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACC 2694
Db 2660 AAAATCGAAATTGACACAACAGGAAGTACTAAACCAAAAAAGGATGACTTGATTCTCACA 2719
Qy 2695 TGTTTCAGAGGATGGGTGAAACAGCTACAGATTGACTACAAAAATCACGAAATCATGACT 2754
Db 2720 TGTTTTAGGGGATGGGTTAAGCAACTGCAAATAGACTACAAAGGAAATGAAATAATGACC 2779
Qy 2755 GCGGCTGCATCGCAAGGACTTACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAAC 2814
Db 2780 GCGGCTGCCTCACAGGGATTGACGCGGAAAGGTGTCTATGCAGTTAGGTACAAAGTTAAC 2839
Qy 2815 GAGAATCCACTCTACTCGCAGACTTCTGAGCACGTGAACGTGTTACTTACACGCACAGAA 2874
Db 2840 GAGAACCCATTGTACGCACCCACCTCAGAGCATGTAAATGTGCTGTTGACCCGGACGGAA 2899
Qy 2875 AAACGCATTGTCTGGAAGACGCTAGCTGGTGATCCCTGGATAAAGACACTTACAGCTAAA 2934



RESULT 8
AAX78130

```
ID   AAX78130 standard; DNA; 11060 BP.
XX
AC   AAX78130;
XX
DT   20-AUG-1999  (first entry)
XX
DE   pSFV link DNA.
XX
KW   Alphavirus; replicon; expression vector; heterologous splice site;
KW   RNA splicing; viral RNA replication; antigen; vaccination; therapy; ss.
XX
OS   Semliki forest virus.
OS   Synthetic.
XX
PN   WO9925859-A1.
XX
PD   27-MAY-1999.
XX
PF   13-NOV-1998;   98WO-CA01065.
XX
PR   14-NOV-1997;   97US-0065793.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Parrington M;
XX
DR   WPI; 1999-385140/32.
XX
PT   New vector (EV) used to express an antigen for vaccination, e.g.
PT   against human immune deficiency virus
XX
PS   Disclosure; Fig 4A-H; 60pp; English.
XX
CC   This invention describes the construction of a novel expression vector
CC   (EV) containing an alphavirus replicon having a heterologous splice site.
CC   The expression vector contains (i) a DNA molecule (I) complementary to
CC   at least part of an alphavirus RNA genome, and (ii) inserted into a
CC   region of (I) that is not essential for its replication, a heterologous
CC   DNA (II), under control of a promoter. (I) includes at least one
CC   heterologous splice site (HSS) to prevent aberrant RNA splicing of the
CC   alphavirus and is the complement of the complete alphavirus genomic
CC   region essential for replication of viral RNA. The expression vector is
CC   used to express (II) in humans or animals, e.g. to express an antigen
CC   for vaccination (against human immune deficiency virus) or to produce a
CC   therapeutically active protein or peptide. Introducing an HSS: (1) makes
CC   it more likely that any splicing will occur at this site, rather than at
CC   a cryptic splice site; (2) restores function of the alphavirus when
CC   eliminated, and (3) may improve transport of RNA from the nucleus.
XX
SQ   Sequence 11060 BP; 2978 A; 2826 C; 2926 G; 2330 T; 0 other;

  Query Match             31.3%;  Score 1565.6;  DB 20;  Length 11060;
  Best Local Similarity   58.3%;  Pred. No. 0;
  Matches 2905;  Conservative    0;  Mismatches 2039;  Indels   36;  Gaps    8;

Qy         31 AGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGTCGTTACAGCGGACGTTT 90
              | | | || |||||  | || ||||||||||| |   ||||||| || |||  | | |||
Db         96 AAAGTGCATGTTGATATTGAGGCTGACAGCCCATTCATCAAGTCTTTGCAGAAGGCATTT 155

Qy         91 CCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACCATGCCAATGCCAGAGCG 150
              ||    || ||| | ||  ||  |||||||||    ||||||||||| ||||||||||
Db        156 CCGTCGTTCGAGGTGGAGTCATTGCAGGTCACACCAAATGACCATGCAAATGCCAGAGCA 215

Qy        151 TTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACCGGGACCAAGTTATCTTG 210
              ||||||||  |||| || ||  | || |||   ||    |||  ||| | | ||||||
Db        216 TTTTCGCACCTGGCTACCAAATTGATCGAGCAGGAGACTGACAAAGACACACTCATCTTG 275

Qy        211 GACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACCGCTATCATTGTATCTGC 270
              || || || ||||||||   |||          || |  |||   || || ||  | |||
Db        276 GATATCGGCAGTGCGCCTTCCAGGAGAATGATGTCTACGCACAAATACCACTGCGTATGC 335

Qy        271 CCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGGTATGCAGAAAGACT------T 324
              ||||||   ||||| |||||||| || || ||  |  | || ||| | | |||

Db        336 CCTATGCGCAGCGCAGAAGACCCCGAAAGGCTCGATAGCTACGCAAAGAAACTGGCAGCG 395

Qy        325 AAGAAAAGTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACCTGCTGGAAGTC 384
                 | |  |    || |   | || ||    ||     | |||||||  |   |||
Db        396 GCCTCCGGGAAGGTGCTGGATAGAGAGATCGCAGGAAAAATCACCGACCTGCAGACCGTC 455

Qy        385 ATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACGCCACGTGTAGG 444
              ||| | || |||||||| ||  |||||  |  | ||  |||| |||||||  |||||||  |
Db        456 ATGGCTACGCCAGACGCTGAATCTCCAACCTTTTGCCTGCATACAGACGTCACGTGTCGT 515

Qy        445 TACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCACCGACATCAATC 504
                  |     || || |||||||| || |||||  || || ||||||||  |||||  |
Db        516 ACGGCAGCCGAAGTGGCCGTATACCAGGACGTGTATGCTGTACATGCACCAACATCGCTG 575

Qy        505 TACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTGACACGACCCCT 564
              |||||  ||||||  | |||||  || || ||    || |||||  || ||||||||  |||||
Db        576 TACCATCAGGCGATGAAAGGTGTCAGAACGGCGTATTGGATTGGGTTTGACACCACCCCG 635

Qy        565 TTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACTGGGCTGACGAG 624
              |||||||    |      | |||||  | || || || |||  ||| ||||||||  ||||||
Db        636 TTTATGTTTGACGCGCTAGCAGGCGCGTATCCAACCTACGCCACAAACTGGGCCGACGAG 695

Qy        625 AGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGGAGAGCAGGCTT 684
                 || ||  | ||  | |||||  || ||  ||     ||     |    ||| | || ||
Db        696 CAGGTGTTACAGGCCAGGAACATAGGACTGTGTGCAGCATCCTTGACTGAGGGAAGACTC 755

Qy        685 GGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGATCATATTCTCG 744
              || ||||| || || ||  | ||||||     |  |||||    | |   |||| ||||||
Db        756 GGCAAACTGTCCATTCTCCGCAAGAAGCAATTGAAACCTTGCGACACAGTCATGTTCTCG 815

Qy        745 GTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGCATCTTCCAAAC 804
              || || || ||| | ||||| ||     |||   ||  |  | |||||||| | ||   |
Db        816 GTAGGATCTACATTGTACACTGAGAGCAGAAAGCTACTGAGGAGCTGGCACTTACCCTCC 875

Qy        805 GTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTGTCAGCTGT 864
              || |||||| ||||||| ||      ||| ||  |||| || |  ||||| ||    |||
Db        876 GTATTCCACCTGAAAGGTAAACAATCCTTTACCTGTAGGTGCGATACCATCGTATCATGT 935

Qy        865 GAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAGTTGAGAAC 924
              ||||||||||  | || || || || ||  |||| || || |||||||||    |
Db        936 GAAGGGTACGTAGTTAAGAAAATCACAATGTGCCCCGGCCTGTACGGTAAAACGGTAGGG 995

Qy        925 TTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATACGCTGCGC 984
              |  ||    ||    ||  |||||  |||  |    |||||    ||||||  || |
Db        996 TACGCCGTGACGTATCACGCGGAGGGATTCCTAGTGTGCAAGACCACAGACACTGTCAAA 1055

Qy        985 GGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCGATCAGATG 1044
              || || || || || ||  |||| || || || || ||  | ||  | || ||||| |||
Db       1056 GGAGAAAGAGTCTCATTCCCTGTATGCACCTACGTCCCCTCAACCATCTGTGATCAAATG 1115

Qy       1045 ACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTATTGGTTGGGCTC 1104
              || || || || || || ||||| |    ||| |||||||| ||  | || || ||  |
Db       1116 ACTGGCATACTAGCGACCGACGTCACACCGGAGGACGCACAGAAGTTGTTAGTGGGATTG 1175

Qy       1105 AACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAATGCAGAACTAT 1164
              || || ||||| || || || || || || ||  |||| |||||||| ||| ||||||||
Db       1176 AATCAGAGGATAGTTGTGAACGGAAGAACACAGCGAAACACTAACACGATGAAGAACTAT 1235

Qy       1165 CTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATCGTGCCGACTTG 1224
              ||  | ||  | || |||   || |||  || ||||||| | ||| |    || ||| |
Db       1236 CTGCTTCCGATTGTGGCCGTCGCATTTAGCAAGTGGGCGAGGGAATACAAGGCAGACCTT 1295

Qy       1225 GACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCTGCTGCTGGGCT 1284
              || || || |||   || || || || ||| |  | ||||||    ||||||  |||||
Db       1296 GATGATGAAAAACCTCTGGGTGTCCGAGAGAGGTCACTTACTTGCTGCTGCTTGTGGGCA 1355

Qy       1285 TTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCTGGTACGCAAACAATTAAGAAA 1344
              || || ||   ||| ||     |||| |||||||| || |  || || |||||   |||
Db       1356 TTTAAAACGAGGAAGATGCACACCATGTACAAGAAACCAGACACCCAGACAATAGTGAAG 1415

Qy       1345 GTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCACGGGCTCGATATG 1404
              || ||| | |  ||| |||| || || || ||  ||||         || ||||  ||
Db       1416 GTGCCTTCAGAGTTTAACTCGTTCGTCATCCCGAGCCTATGGTCTACAGGCCTCGCAATC 1475
```


Qy 1405 GGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACTGTCAAACCCGCACCGGCTATT 1464
Db 1476 CCAGTCAGATCACGCATTAAGATGCTTTTGGCCAAGAAGACCAAGC---GAGAGTTAATA 1532
Qy 1465 ACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAGAAGTGGCTGCA 1524
Db 1533 CCTGTTCTCGACGCGTCGTCAGCCAGGGATGCTGAACAAGAGGAGAAGGAGAGGTTGGAG 1592
Qy 1525 GCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAAAAGAGACCGTA 1584
Db 1593 GCCGAGCTGACTAGAGAAGCCTTACCACCCCTCGTCCCCATCGCGCCGGCGGAGACGGGA 1652
Qy 1585 GAGGCAGAAGTAGACCTCATTATGCAAGAG------GCAGGAGCAGGTAGCGTGGAGACA 1638
Db 1653 GTCGTCGACGTCGACGTTGAAGAACTAGAGTATCACGCAGGTGCAGGGGTCGTGGAAACA 1712
Qy 1639 CCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAGATTGGGTCTTACGCT 1698
Db 1713 CCTCGCAGCGCGTTGAAAGTCACCGCACAGCCGAACGACGTACTACTAGGAAATTACGTA 1772
Qy 1699 ATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGTATCCACCCATTGGCG 1758
Db 1773 GTTCTGTCCCCGCAGACCGTGCTCAAGAGCTCCAAGTTGGCCCCCGTGCACCCTCTAGCA 1832
Qy 1759 GAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATACAAAGTCGAGCCATAC 1818
Db 1833 GAGCAGGTGAAAATAATAACACATAACGGGAGGGCCGGCGGTTACCAGGTCGACGGATAT 1892
Qy 1819 CACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAAGACTTCCAGGCATTG 1878
Db 1893 GACGGCAGGGTCCTACTACCATGTGGATCGGCCATTCCGGTCCCTGAGTTTCAAGCTTTG 1952
Qy 1879 AGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAACAGATACCTGCACCAC 1938
Db 1953 AGCGAGAGCGCCACTATGGTGTACAACGAAAGGGAGTTCGTCAACAGGAAACTATACCAT 2012
Qy 1939 ATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTATAAGACTGTAAAGACT 1998
Db 2013 ATTGCCGTTCACGGACCGTCGCTGAACACCGACGAGGAGAACTACGAGAAAGTCAGAGCT 2072
Qy 1999 CAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAGTGTGTTAAGCGAGAA 2058
Db 2073 GAAAGAACTGACGCCGAGTACGTGTTCGACGTAGATAAAAAATGCTGCGTCAAGAGAGAG 2132
Qy 2059 GACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCATTTCACGAGTTTGCG 2118
Db 2133 GAAGCGTCGGGTTTGGTGTTGGTGGGAGAGCTAACCAACCCCCCGTTCCATGAATTCGCC 2192
Qy 2119 TACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCAACCATCGGAGTCTAT 2178
Db 2193 TACGAAGGGCTGAAGATCAGGCCGTCGGCACCATATAAGACTACAGTAGTAGGAGTCTTT 2252
Qy 2179 GGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTGACTAAGAAAGATCTG 2238
Db 2253 GGGGTTCCGGGATCAGGCAAGTCTGCTATTATTAAGAGCCTCGTGACCAAACACGATCTG 2312
Qy 2239 GTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGATGTAAGGAGGATGAGA 2298
Db 2313 GTCACCAGCGGCAAGAAGGAGAACTGCCAGGAAATAGTTAACGACGTGAAGAAGCACCGC 2372
Qy 2299 CGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAATGGGGTTAAGCACCCC 2358
Db 2373 GGGAAGGGGACAAGTAGGGAAAACAGTGACTCCATCCTGCTAAACGGGTGTCGTCGTGCC 2432
Qy 2359 GTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGGACGCTGCTGGCACTG 2418
Db 2433 GTGGACATCCTATATGTGGACGAGGCTTTCGCTTGCCATTCCGGTACTCTGCTGGCCCTA 2492
Qy 2419 ATTGCCATCGTCAAACCT---AAGAAAGTGGTATTGTGCGGGGACCCAAAACAATGCGGC 2475
Db 2493 ATTGCTCTTGTTAAACCTCGGAGCAAAGTGGTGTTATGCGGAGACCCCAAGCAATGCGGA 2552
Qy 2476 TTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGCACTGAAGTGTAC 2535
Db 2553 TTCTTCAATATGATGCAGCTTAAGGTGAACTTCAACCACAACATCTGCACTGAAGTATGT 2612
Qy 2536 CATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATCGTCTCCACGCTCTTC 2595
Db 2613 CATAAAAGTATATCCAGACGTTGCACGCGTCCAGTCACGGCCATCGTGTCTACGTTGCAC 2672
Qy 2596 TACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATCATCATAGATACCACA 2655
Db 2673 TACGGAGGCAAGATGCGCACGACCAACCCGTGCAACAAACCCATAATCATAGACACCACA 2732
Qy 2656 GGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGAGGATGGGTGAAA 2715
Db 2733 GGACAGACCAAGCCCAAGCCAGGAGACATCGTGTTAACATGCTTCCGAGGCTGGGCAAAG 2792
Qy 2716 CAGCTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCTGCATCGCAAGGACTT 2775
Db 2793 CAGCTGCAGTTGGACTACCGTGGACACGAAGTCATGACAGCAGCAGCATCTCAGGGCCTC 2852
Qy 2776 ACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCACTCTACTCGCAG 2835
Db 2853 ACCCGCAAAGGGGTATACGCCGTAAGGCAGAAGGTGAATGAAAATCCCTTGTATGCCCCT 2912
Qy 2836 ACTTCTGAGCACGTGAACGTGTTACTTACACGCACAGAAAAACGCATTGTCTGGAAGACG 2895
Db 2913 GCGTCGGAGCACGTGAATGTACTGCTGACGCGCACTGAGGATAGGCTGGTGTGGAAAACG 2972
Qy 2896 CTAGCTGGTGATCCCTGGATAAAGACACTTACAGCTAAATATCCCGGGGATTTCACGGCT 2955
Db 2973 CTGGCCGGCGATCCCTGGATTAAGGTCCTATCAAACATTCCACAGGGTAACTTTACGGCC 3032
Qy 2956 TCATTGGACGACTGGCAGCGCGAACACGACGCCATTATGGCACGCGTTCTTGATAAGCCG 3015
Db 3033 ACATTGGAAGAATGGCAAGAAGAACACGACAAAATAATGAAGGTGATTGAAGGACCGGCT 3092
Qy 3016 CAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAGGCTTTAGAGCCA 3075
Db 3093 GCGCCTGTGGACGCGTTCCAGAACAAAGCGAACGTGTGTTGGGCGAAAAGCCTGGTGCCT 3152
Qy 3076 GTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAGAC---GTTGCACCCA 3132
Db 3153 GTCCTGGACACTGCCGGAATCAGATTGACAGCAGAGGAGTGGAGCACCATAATTACAGCA 3212
Qy 3133 TTCAAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTTTGCACCAGGTTC 3192
Db 3213 TTTAAGGAGGACAGAGCTTACTCTCCAGTGGTGGCCTTGAATGAAATTTGCACCAAGTAC 3272
Qy 3193 TTTGGAGTAGACCTGGACAGTGGGTTATTTTCCGCTCCTACCGTCGCACTTACTTACAGG 3252
Db 3273 TATGGAGTTGACCTGGACAGTGGCCTGTTTTCTGCCCCGAAGGTGTCCCTGTATTACGAG 3332
Qy 3253 GATCAGCACTGGGATAACTCGCCAGGGAAGAACATGTATGGGCTTAATAGAGAGGTAGCA 3312
Db 3333 AACAACCACTGGGATAACAGACCTGGTGGAAGGATGTATGGATTCAATGCCGCAACAGCT 3392
Qy 3313 AAGGAGTTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACAGGCAGGGTAGCT 3372
Db 3393 GCCAGGCTGGAAGCTAGACATACCTTCCTGAAGGGGCAGTGGCATACGGGCAAGCAGGCA 3452
Qy 3373 GATATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTGGTTCCATTAAAT 3432
Db 3453 GTTATCGCAGAAAGAAAAATCCAACCGCTTTCTGTGCTGGACAATGTAATTCCTATCAAC 3512
Qy 3433 CGCCGGTTGCCCCACTCGTTGATCGTTGACCACAAAGGACAGGGTACAACTGATCACAGC 3492
Db 3513 CGCAGGCTGCCGCACGCCCTGGTGGCTGAGTACAAGACGGTTAAAGGCAGTAGGGTTGAG 3572
Qy 3493 GGATTCCTATCTAAGATGAAGGGCAAATCTGTGTTGGTGATCGGCGATCCTATCAGCATT 3552
Db 3573 TGGCTGGTCAATAAAGTAAGAGGGTACCACGTCCTGCTGGTGAGTGAGTACAACCTGGCT 3632
Qy 3553 CCAGGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATAC---------CATCAGG 3603

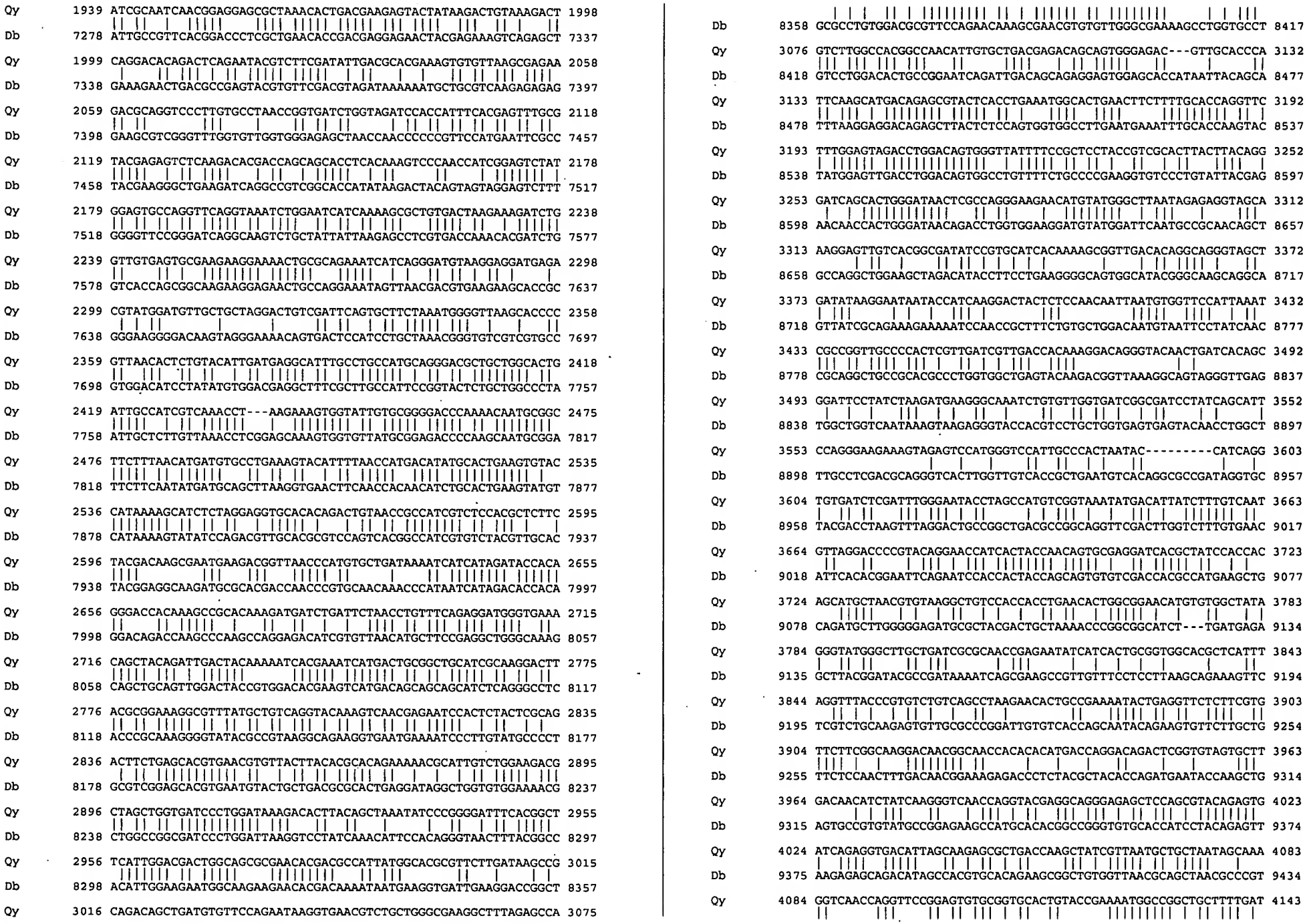


RESULT 9
AAC55473
ID AAC55473 standard; DNA; 12464 BP.
XX
AC AAC55473;
XX
DT 11-JAN-2001 (first entry)
XX
DE Destination vector pDEST9 nucleotide sequence.
XX
KW Bacteriophage lambda; att; recombination site; attB; attP; attR; attL;
KW mutant; recombinational cloning; entry vector; destination vector;
KW gene product targeting; fusion tag cleavage; ds.
XX
OS Bacteriophage lambda.
OS Synthetic.
XX
PN WO200052027-A1.
XX
PD 08-SEP-2000.
XX
PF 02-MAR-2000; 2000WO-US05432.
XX
PR 02-MAR-1999; 99US-0122389.
PR 23-MAR-1999; 99US-0126049.
PR 28-MAY-1999; 99US-0136744.
XX
PA (LIFE-) LIFE TECHNOLOGIES INC.
XX
PI Hartley JL, Brasch MA, Temple GF, Cheo D;
XX
DR WPI; 2000-543948/49.
XX
PT Isolated nucleic acid molecules encoding an attB1, attB2, attP1, attP2,
PT attL1, attL2, attR1, and attR2 nucleotide sequence useful for the
PT recombinational cloning of polypeptides -
XX
PS Disclosure; Fig 29; 459pp; English.
XX
CC The present invention describes isolated nucleic acid molecules (I)
CC encoding an attB1, attB2, attP1, attP2, attL1, attL2, attR1, and attR2
CC nucleotide sequence. Also described are: (1) an isolated nucleic acid
CC molecule (II) comprising one or more att recombination sites comprising
CC at least one mutation in its core region that increases the specificity
CC of interaction between the recombination site and a second att
CC recombination site; and (2) an isolated nucleic acid molecule (III)
CC comprising one or more mutated att recombination sites comprising at
CC least one mutation in its core region that enhances the efficiency of
CC recombination between a first nucleic acid molecule comprising the
CC mutated att recombination site and a second nucleic acid molecule

CC comprising a second recombination site that interacts with the mutated
CC att recombination site. (I), (II), (III), primers, vectors and methods
CC from the present invention are used for the recombinational cloning of
CC nucleic acid molecules. They can be used for changing vectors, targeting
CC gene products to intracellular locations, cleaving fusion tags from
CC desired proteins, operably linking nucleic acid molecules of interest to
CC regulatory genetic sequences, constructing genes for fusion proteins,
CC changing copy number, changing replicons, cloning into phages and
CC cloning. (I), (II), (III), host cells and vectors can be used in the
CC production of polypeptides and antibodies. The present sequence is
CC used in the exemplification of the present invention.
XX
SQ Sequence 12464 BP; 3328 A; 3120 C; 3257 G; 2759 T; 0 other;

Query Match 31.3%; Score 1565.6; DB 21; Length 12464;
Best Local Similarity 58.3%; Pred. No. 0;
Matches 2905; Conservative 0; Mismatches 2039; Indels 36; Gaps 8;

```
Qy        31 AGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGTCGTTACAGCGGACGTTT 90
Db      5361 AAAGTGCATGTTGATATTGAGGCTGACAGCCCATTCATCAAGTCTTTGCAGAAGGCATTT 5420

Qy        91 CCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACCATGCCAATGCCAGAGCG 150
Db      5421 CCGTCGTTCGAGGTGGAGTCATTGCAGGTCACACCAAATGACCATGCAAATGCCAGAGCA 5480

Qy       151 TTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACCGGGACCAAGTTATCTTG 210
Db      5481 TTTTCGCACCTGGCTACCAAATTGATCGAGCAGGAGACTGACAAAGACACACTCATCTTG 5540

Qy       211 GACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACCGCTATCATTGTATCTGC 270
Db      5541 GATATCGGCAGTGCGCCTTCCAGGAGAATGATGTCTACGCACAAATACCACTGCGTATGC 5600

Qy       271 CCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGGTATGCAGAAAGACT------T 324
Db      5601 CCTATGCGCAGCGCAGAAGACCCCGAAAGGCTCGATAGCTACGCAAAGAAACTGGCAGCG 5660

Qy       325 AAGAAAAGTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACCTGCTGGAAGTC 384
Db      5661 GCCTCCGGGAAGGTGCTGGATAGAGAGATCGCAGGAAAAATCACCGACCTGCAGACCGTC 5720

Qy       385 ATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACGCCACGTGTAGG 444
Db      5721 ATGGCTACGCCAGACGCTGAATCTCCTACCTTTTGCCTGCATACAGACGTCACGTGTCGT 5780

Qy       445 TACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCACCGACATCAATC 504
Db      5781 ACGGCAGCCGAAGTGGCCGTATACCAGGACGTGTATGCTGTACATGCACCAACATCGCTG 5840

Qy       505 TACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTGACACGACCCCT 564
Db      5841 TACCATCAGGCGATGAAAGGTGTCAGAACGGCGTATTGGATTGGGTTTGACACCACCCCG 5900

Qy       565 TTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACTGGGCTGACGAG 624
Db      5901 TTTATGTTTGACGCGCTAGCAGGCGCGTATCCAACCTACGCCACAAACTGGGCCGACGAG 5960

Qy       625 AGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGGAGAGCAGGCTT 684
Db      5961 CAGGTGTTACAGGCCAGGAACATAGGACTGTGTGCAGCATCCTTGACTGAGGGAAGACTC 6020

Qy       685 GGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGATCATATTCTCG 744
Db      6021 GGCAAACTGTCCATTCTCCGCAAGAAGCAATTGAAACCTTGCGACACAGTCATGTTCTCG 6080

Qy       745 GTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGCATCTTCCAAAC 804
Db      6081 GTAGGATCTACATTGTACACTGAGAGCAGAAAGCTACTGAGGAGCTGGCACTTACCCTCC 6140

Qy       805 GTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTGTCAGCTGT 864
Db      6141 GTATTCCACCTGAAAGGTAAACAATCCTTTACCTGTAGGTGCGATACCATCGTATCATGT 6200

Qy       865 GAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAGTTGAGAAC 924
Db      6201 GAAGGGTACGTAGTTAAGAAAATCACTATGTGCCCCGGCCTGTACGGTAAAACGGTAGGG 6260

Qy       925 TTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATACGCTGCGC 984
Db      6261 TACGCCGTGACGTATCACGCGGAGGGATTCCTAGTGTGCAAGACCACAGACACTGTCAAA 6320

Qy       985 GGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCGATCAGATG 1044
Db      6321 GGAGAAAGAGTCTCATTCCCTGTATGCACCTACGTCCCCTCAACCATCTGTGATCAAATG 6380

Qy      1045 ACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTATTGGTTGGGCTC 1104
Db      6381 ACTGGCATACTAGCGACCGACGTCACACCGGAGGACGCACAGAAGTTGTTAGTGGGATTG 6440

Qy      1105 AACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAATGCAGAACTAT 1164
Db      6441 AATCAGAGGATAGTTGTGAACGGAAGAACACAGCGAAACACTAACACGATGAAGAACTAT 6500

Qy      1165 CTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATCGTGCCGACTTG 1224
Db      6501 CTGCTTCCGATTGTGGCCGTCGCATTTAGCAAGTGGGCGAGGGAATACAAGGCAGACCTT 6560

Qy      1225 GACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCTGCTGCTGGGCT 1284
Db      6561 GATGATGAAAAACCTCTGGGTGTCCGAGAGAGGTCACTTACTTGCTGCTGCTTGTGGGCA 6620

Qy      1285 TTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCTGGTACGCAAACAATTAAGAAA 1344
Db      6621 TTTAAAACGAGGAAGATGCACACCATGTACAAGAAACCAGACACCCAGACAATAGTGAAG 6680

Qy      1345 GTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCACGGGCTCGATATG 1404
Db      6681 GTGCCTTCAGAGTTTAACTCGTTCGTCATCCCGAGCCTATGGTCTACAGGCCTCGCAATC 6740

Qy      1405 GGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACTGTCAAACCCGCACCGGCTATT 1464
Db      6741 CCAGTCAGATCACGCATTAAGATGCTTTTGGCCAAGAAGACCAAGC---GAGAGTTAATA 6797

Qy      1465 ACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAGAAGTGGCTGCA 1524
Db      6798 CCTGTTCTCGACGCGTCGTCAGCCAGGGATGCTGAACAAGAGGAGAAGGAGAGGTTGGAG 6857

Qy      1525 GCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAAAAGAGACCGTA 1584
Db      6858 GCCGAGCTGACTAGAGAAGCCTTACCACCCCTCGTCCCCATCGCGCCGGCGGAGACGGGA 6917

Qy      1585 GAGGCAGAAGTAGACCTCATTATGCAAGAG------GCAGGAGCAGGTAGCGTGGAGACA 1638
Db      6918 GTCGTCGACGTCGACGTTGAAGAACTAGAGTATCACGCAGGTGCAGGGGTCGTGGAAACA 6977

Qy      1639 CCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAGATTGGGTCTTACGCT 1698
Db      6978 CCTCGCAGCGCGTTGAAAGTCACCGCACAGCCGAACGACGTACTACTAGGAAATTACGTA 7037

Qy      1699 ATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGTATCCACCCATTGGCG 1758
Db      7038 GTTCTGTCCCCGCAGACCGTGCTCAAGAGCTCCAAGTTGGCCCCCGTGCACCCTCTAGCA 7097

Qy      1759 GAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATACAAAGTCGAGCCATAC 1818
Db      7098 GAGCAGGTGAAAATAATAACACATAACGGGAGGGCCGGCGGTTACCAGGTCGACGGATAT 7157

Qy      1819 CACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAAGACTTCCAGGCATTG 1878
Db      7158 GACGGCAGGGTCCTACTACCATGTGGATCGGCCATTCCGGTCCCTGAGTTTCAGGCTTTG 7217

Qy      1879 AGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAACAGATACCTGCACCAC 1938
Db      7218 AGCGAGAGCGCCACTATGGTGTACAACGAAAGGGAGTTCGTCAACAGGAAACTATACCAT 7277
```

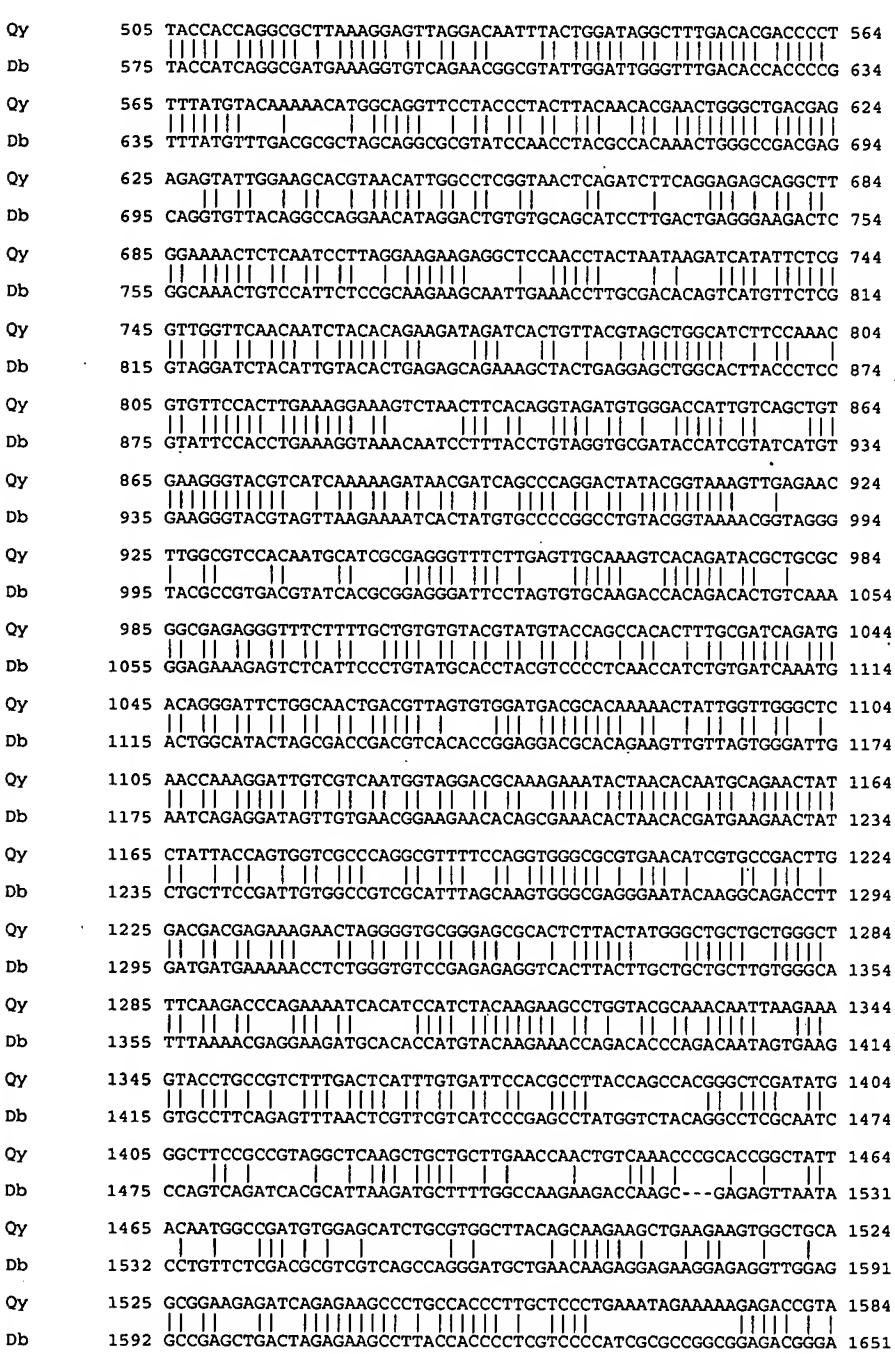


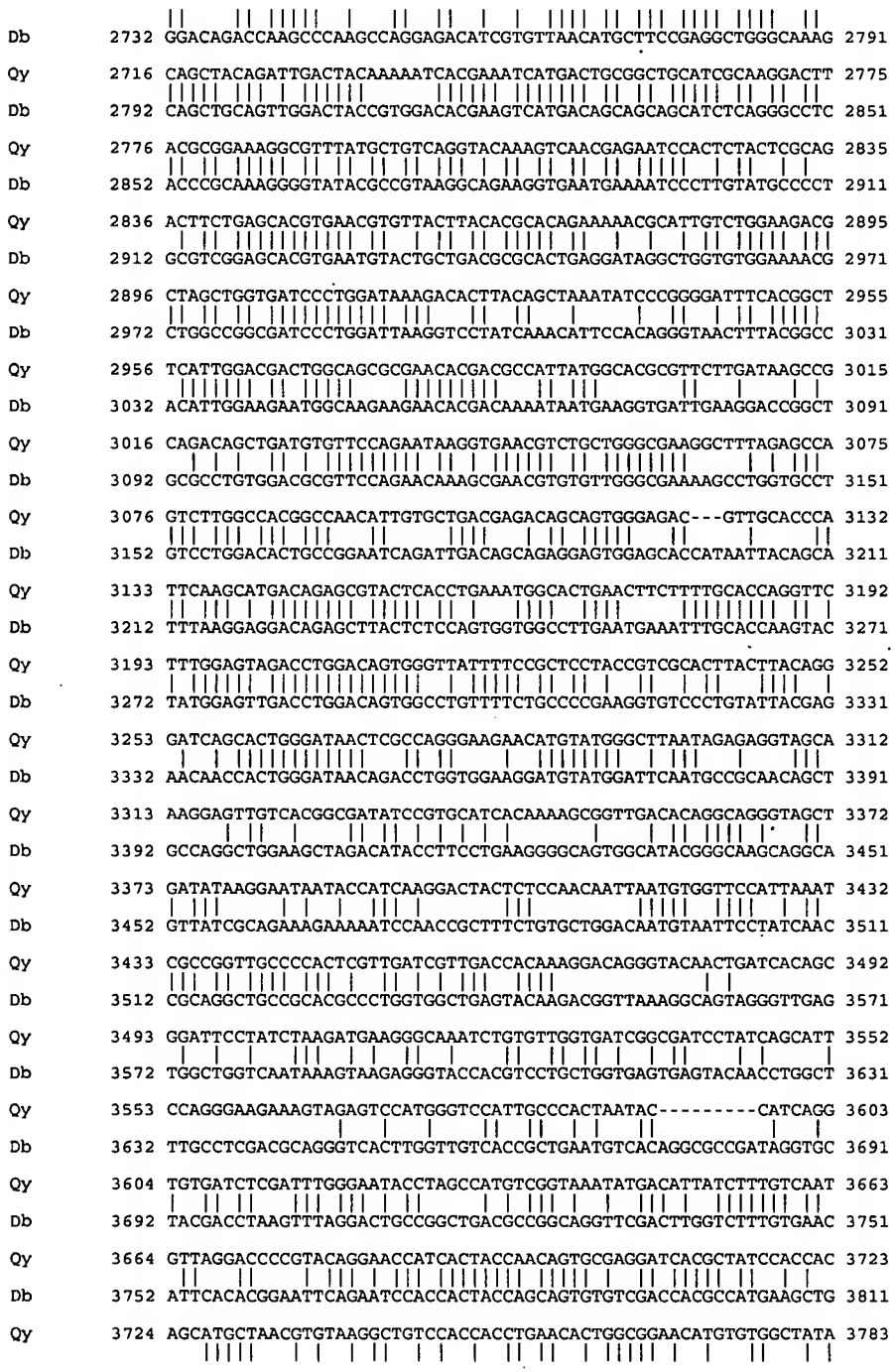
```
Db       9435 GGAACTGTAGGGGATGGCGTATGCAGGGCCGTGGCGAAGAAATGGCCGTCAGCCTTTAAG 9494

Qy       4144 AGACAGCCAATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACCG---CTCATCATA 4200

Db       9495 GGAGCAGCAACACCAGTGGGCACAATTAAAACAGTCATGTGCGGCTCGTACCCCGTCATC 9554

Qy       4201 CATGCTGTAGGACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTAAGCTCGCA 4260

Db       9555 CACGCTGTAGCGCCTAATTTCTCTGCCACGACTGAAGCGGAAGGGGACCGCGAATTGGCC 9614

Qy       4261 GCTGCCTACATGAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTA 4320

Db       9615 GCTGTCTACCGGGCAGTGGCCGCCGAAGTAAACAGACTGTCACTGAGCAGCGTAGCCATC 9674

Qy       4321 CCGCTACTGTCAACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCAT 4380

Db       9675 CCGCTGCTGTCCACAGGAGTGTTCAGCGGCGGAAGAGATAGGCTGCAGCAATCCCTCAAC 9734

Qy       4381 CACCTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAA 4440

Db       9735 CATCTATTCACAGCAATGGACGCCACGGACGCTGACGTGACCATCTACTGCAGAGACAAA 9794

Qy       4441 CAATGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATTCTGGAT 4500

Db       9795 AGTTGGGAGAAGAAAATCCAGGAAGCCATTGACATGAGGACGGCTGTGGAGTTGCTCAAT 9854

Qy       4501 GATGACAAGCCAGTAGACATTGACTTGGTCAGGGTCCACCCAAACAGCTCTTTGGCAGGC 4560

Db       9855 GATGACGTGGAGCTGACCACAGACTTGGTGAGAGTGCACCCGGACAGCAGCCTGGTGGGT 9914

Qy       4561 AGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTGGAAGGTACACGATTC 4620

Db       9915 CGTAAGGGCTACAGTACCACTGACGGGTCGCTGTACTCGTACTTTGAAGGTACGAAATTC 9974

Qy       4621 CATCAGACCGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCCAACAAATCTGAGGCT 4680

Db       9975 AACCAGGCTGCTATTGATATGGCAGAGATACTGACGTTGTGGCCCAGACTGCAAGAGGCA 10034

Qy       4681 AATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGC 4740

Db       10035 AACGAACAGATATGCCTATACGCGCTGGGCGAAACAATGGACAACATCAGATCCAAATGT 10094

Qy       4741 CCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTAC 4800

Db       10095 CCGGTGAACGATTCCGATTCATCAACACCTCCCAGGACAGTGCCCTGCCTGTGCCGCTAC 10154

Qy       4801 GCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTA 4860

Db       10155 GCAATGACAGCAGAACGGATCGCCCGCCTTAGGTCACACCAAGTTAAAAGCATGGTGGTT 10214

Qy       4861 TGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGC 4920

Db       10215 TGCTCATCTTTTCCCCTCCCGAAATACCATGTAGATGGGGTGCAGAAGGTAAAGTGCGAG 10274

Qy       4921 AAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGCGGAA 4980

Db       10275 AAGGTTCTCCTGTTCGACCCGACGGTACCTTCAGTGGTTAGTCCGCGGAAGTATGCCGCA 10334
```

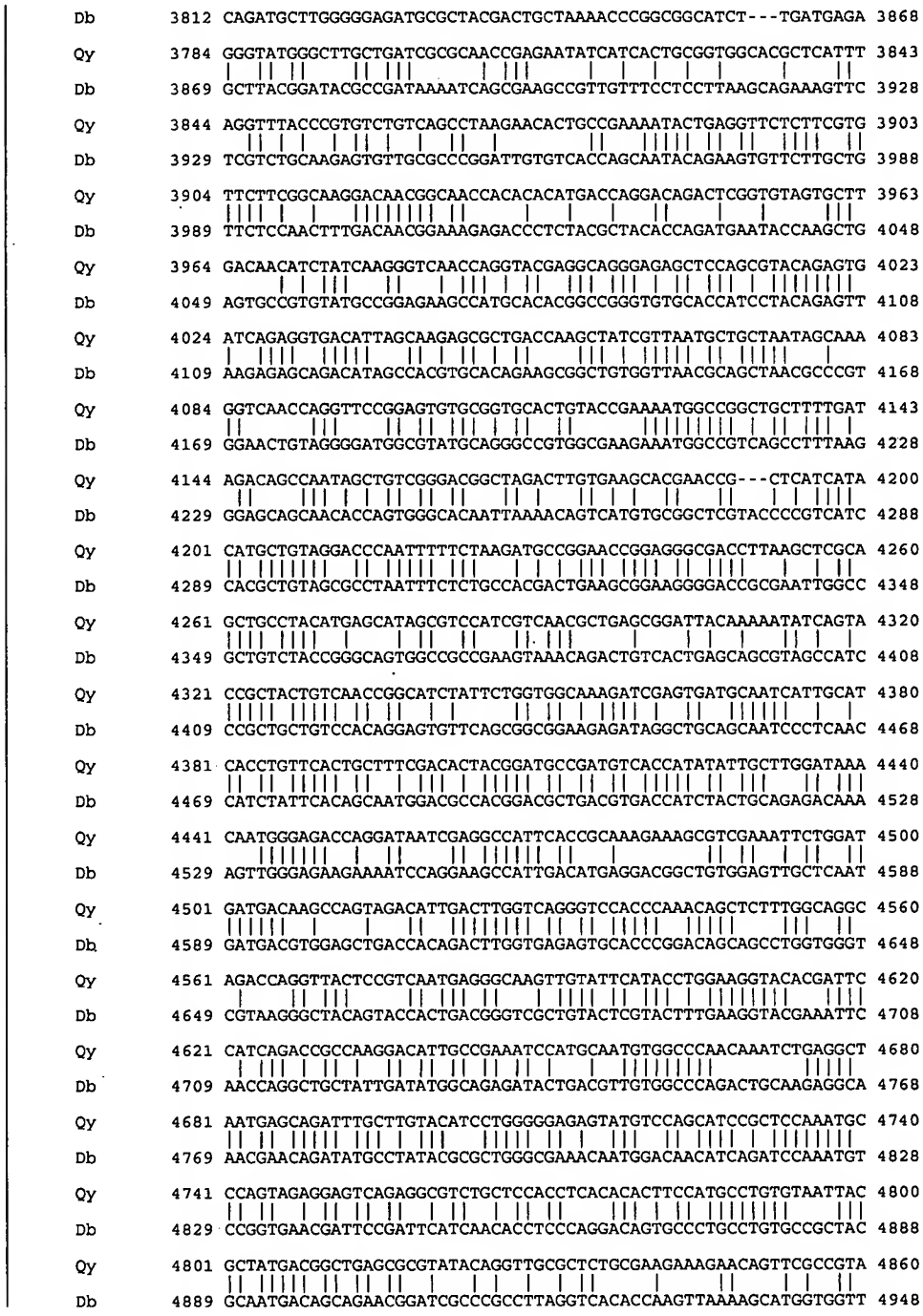
RESULT 10
ACC45139
ID ACC45139 standard; DNA; 7399 BP.
XX
AC ACC45139;
XX
DT 11-JUN-2003 (first entry)
XX
DE Alphavirus replication unit NS1-NS4 gene related sequence.
XX
KW Vaccine; alpha virus replication unit; NS1-NS4; gene; ds.
XX
OS Alphavirus.
XX
PN CN1367020-A.
XX
PD 04-SEP-2002.
XX
PF 19-OCT-2001; 2001CN-0128979.
XX
PR 19-OCT-2001; 2001CN-0128979.
XX
PA (UYHU-) UNIV HUAXI HOSPITAL SICHUAN.
XX
PI Zhao L, Qin S;
XX
DR WPI; 2003-168348/17.
XX
PT Nucleic acid vaccine capable of being high-effectively and safely
PT expressed comprises a conventional nucleic acid vaccine plasmid and
PT alpha virus replication unit NS1-NS4 -
XX
PS Example; Fig 1; 14pp; Chinese.
XX
CC The present invention describes a nucleic acid vaccine which can be
CC effectively safely expressed. The nucleic acid vaccine plasmid forming
CC the nucleic acid vaccine is formed from a conventional nucleic acid
CC vaccine plasmid and an alphavirus replication unit NS1-NS4 gene
CC recombination. After having been injected into body, under the action
CC of a CMV promoter, the nucleic acid vaccine can transcribe the NS1-NS4
CC replication unit and RNA of its downstream target gene with high
CC replication, and it possess high expression efficiency of its target
CC antigen, and after its transcription in the cell and expression level
CC has reach a certain high level, it can induce the cell transfected by
CC performing a death process to form a self-destroying mechanism so as to
CC effectively solve the safety problem of nucleic acid vaccines. The
CC present sequence represents an alphavirus replication unit NS1-NS4 gene
CC related nucleotide sequence, which is used in an example from the
CC present invention.
XX
SQ Sequence 7399 BP; 1966 A; 1915 C; 2025 G; 1493 T; 0 other;

Query Match 31.3%; Score 1564; DB 25; Length 7399;
Best Local Similarity 58.3%; Pred. No. 0;
Matches 2904; Conservative 0; Mismatches 2040; Indels 36; Gaps 8;

```
Qy         31 AGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGTCGTTACAGCGGACGTTT 90

Db         95 AAAGTGCATGTTGATATTGAGGCTGACAGCCCATTCATCAAGTCTTTGCAGAAGGCATTT 154

Qy         91 CCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACCATGCCAATGCCAGAGCG 150

Db        155 CCGTCGTTCGAGGTGGAGTCATTGCAGGTCACACCAAATGACCATGCAAATGCCAGAGCA 214

Qy        151 TTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACCGGGACCAAGTTATCTTG 210

Db        215 TTTTCGCACCTGGCTACCAAATTGATCGAGCAGGAGACTGACAAAGACACACTCATCTTG 274

Qy        211 GACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACCGCTATCATTGTATCTGC 270

Db        275 GATATCGGCAGTGCGCCTTCCAGGAGAATGATGTCTACGCACAAATACCACTGCGTATGC 334

Qy        271 CCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGGTATGCAGAAAGACT------T 324

Db        335 CCTATGCGCAGCGCAGAAGACCCCGAAAGGCTCGATAGCTACGCAAAGAAACTGGCAGCG 394

Qy        325 AAGAAAAGTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACCTGCTGGAAGTC 384

Db        395 GCCTCCGGGAAGGTGCTGGATAGAGAGATCGCAGGAAAAATCACCGACCTGCAGACCGTC 454

Qy        385 ATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACGCCACGTGTAGG 444

Db        455 ATGGCTACGCCAGACGCTGAATCTCCTACCTTTTGCCTGCATACAGACGTCACGTGTCGT 514

Qy        445 TACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCACCGACATCAATC 504

Db        515 ACGGCAGCCGAAGTGGCCGTATACCAGGACGTGTATGCTGTACATGCACCAACATCGCTG 574
```


Qy 505 TACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTGACACGACCCCT 564
Db 575 TACCATCAGGCGATGAAAGGTGTCAGAACGGCGTATTGGATTGGGTTTGACACCACCCCG 634
Qy 565 TTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACTGGGCTGACGAG 624
Db 635 TTTATGTTTGACGCGCTAGCAGGCGCGTATCCAACCTACGCCACAAACTGGGCCGACGAG 694
Qy 625 AGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGGAGAGCAGGCTT 684
Db 695 CAGGTGTTACAGGCCAGGAACATAGGACTGTGTGCAGCATCCTTGACTGAGGGAAGACTC 754
Qy 685 GGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGATCATATTCTCG 744
Db 755 GGCAAACTGTCCATTCTCCGCAAGAAGCAATTGAAACCTTGCGACACAGTCATGTTCTCG 814
Qy 745 GTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGCATCTTCCAAAC 804
Db 815 GTAGGATCTACATTGTACACTGAGAGCAGAAAGCTACTGAGGAGCTGGCACTTACCCTCC 874
Qy 805 GTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTGTCAGCTGT 864
Db 875 GTATTCCACCTGAAAGGTAAACAATCCTTTACCTGTAGGTGCGATACCATCGTATCATGT 934
Qy 865 GAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAGTTGAGAAC 924
Db 935 GAAGGGTACGTAGTTAAGAAAATCACTATGTGCCCCGGCCTGTACGGTAAAACGGTAGGG 994
Qy 925 TTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATACGCTGCGC 984
Db 995 TACGCCGTGACGTATCACGCGGAGGGATTCCTAGTGTGCAAGACCACAGACACTGTCAAA 1054
Qy 985 GGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCGATCAGATG 1044
Db 1055 GGAGAAAGAGTCTCATTCCCTGTATGCACCTACGTCCCCTCAACCATCTGTGATCAAATG 1114
Qy 1045 ACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTATTGGTTGGGCTC 1104
Db 1115 ACTGGCATACTAGCGACCGACGTCACACCGGAGGACGCACAGAAGTTGTTAGTGGGATTG 1174
Qy 1105 AACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAATGCAGAACTAT 1164
Db 1175 AATCAGAGGATAGTTGTGAACGGAAGAACACAGCGAAACACTAACACGATGAAGAACTAT 1234
Qy 1165 CTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATCGTGCCGACTTG 1224
Db 1235 CTGCTTCCGATTGTGGCCGTCGCATTTAGCAAGTGGGCGAGGGAATACAAGGCAGACCTT 1294
Qy 1225 GACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCTGCTGCTGGGCT 1284
Db 1295 GATGATGAAAAACCTCTGGGTGTCCGAGAGAGGTCACTTACTTGCTGCTGCTTGTGGGCA 1354
Qy 1285 TTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCTGGTACGCAAACAATTAAGAAA 1344
Db 1355 TTTAAAACGAGGAAGATGCACACCATGTACAAGAAACCAGACACCCAGACAATAGTGAAG 1414
Qy 1345 GTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCACGGGCTCGATATG 1404
Db 1415 GTGCCTTCAGAGTTTAACTCGTTCGTCATCCCGAGCCTATGGTCTACAGGCCTCGCAATC 1474
Qy 1405 GGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACTGTCAAACCCGCACCGGCTATT 1464
Db 1475 CCAGTCAGATCACGCATTAAGATGCTTTTGGCCAAGAAGACCAAGC---GAGAGTTAATA 1531
Qy 1465 ACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAGAAGTGGCTGCA 1524
Db 1532 CCTGTTCTCGACGCGTCGTCAGCCAGGGATGCTGAACAAGAGGAGAAGGAGAGGTTGGAG 1591
Qy 1525 GCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAAAAGAGACCGTA 1584
Db 1592 GCCGAGCTGACTAGAGAAGCCTTACCACCCCTCGTCCCCATCGCGCCGGCGGAGACGGGA 1651
Qy 1585 GAGGCAGAAGTAGACCTCATTATGCAAGAG------GCAGGAGCAGGTAGCGTGGAGACA 1638
Db 1652 GTCGTCGACGTCGACGTTGAAGAACTAGAGTATCACGCAGGTGCAGGGGTCGTGGAAACA 1711
Qy 1639 CCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAGATTGGGTCTTACGCT 1698
Db 1712 CCTCGCAGCGCGTTGAAAGTCACCGCACAGCCGAACGACGTACTACTAGGAAATTACGTA 1771
Qy 1699 ATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGTATCCACCCATTGGCG 1758
Db 1772 GTTCTGTCCCCGCAGACCGTGCTCAAGAGCTCCAAGTTGGCCCCCGTGCACCCTCTAGCA 1831
Qy 1759 GAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATACAAAGTCGAGCCATAC 1818
Db 1832 GAGCAGGTGAAAATAATAACACATAACGGGAGGGCCGGCGGTTACCAGGTCGACGGATAT 1891
Qy 1819 CACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAAGACTTCCAGGCATTG 1878
Db 1892 GACGGCAGGGTCCTACTACCATGTGGATCGGCCATTCCGGTCCCTGAGTTTCAAGCTTTG 1951
Qy 1879 AGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAACAGATACCTGCACCAC 1938
Db 1952 AGCGAGAGCGCCACTATGGTGTACAACGAAAGGGAGTTCGTCAACAGGAAACTATACCAT 2011
Qy 1939 ATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTATAAGACTGTAAAGACT 1998
Db 2012 ATTGCCGTTCACGGACCGTCGCTGAACACCGACGAGGAGAACTACGAGAAAGTCAGAGCT 2071
Qy 1999 CAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAGTGTGTTAAGCGAGAA 2058
Db 2072 GAAAGAACTGACGCCGAGTACGTGTTCGACGTAGATAAAAAATGCTGCGTCAAGAGAGAG 2131
Qy 2059 GACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCATTTCACGAGTTTGCG 2118
Db 2132 GAAGCGTCGGGTTTGGTGTTGGTGGGAGAGCTAACCAACCCCCCGTTCCATGAATTCGCC 2191
Qy 2119 TACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCAACCATCGGAGTCTAT 2178
Db 2192 TACGAAGGGCTGAAGATCAGGCCGTCGGCACCATATAAGACTACAGTAGTAGGAGTCTTT 2251
Qy 2179 GGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTGACTAAGAAAGATCTG 2238
Db 2252 GGGGTTCCGGGATCAGGCAAGTCTGCTATTATTAAGAGCCTCGTGACCAAACACGATCTG 2311
Qy 2239 GTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGATGTAAGGAGGATGAGA 2298
Db 2312 GTCACCAGCGGCAAGAAGGAGAACTGCCAGGAAATAGTTAACGACGTGAAGAAGCACCGC 2371
Qy 2299 CGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAATGGGGTTAAGCACCCC 2358
Db 2372 GGGAAGGGGACAAGTAGGGAAAACAGTGACTCCATCCTGCTAAACGGGTGTCGTCGTGCC 2431
Qy 2359 GTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGGACGCTGCTGGCACTG 2418
Db 2432 GTGGACATCCTATATGTGGACGAGGCTTTCGCTTGCCATTCCGGTACTCTGCTGGCCCTA 2491
Qy 2419 ATTGCCATCGTCAAACCT---AAGAAAGTGGTATTGTGCGGGGACCCAAAACAATGCGGC 2475
Db 2492 ATTGCTCTTGTTAAACCTCGGAGCAAAGTGGTGTTATGCGGAGACCCCAAGCAATGCGGA 2551
Qy 2476 TTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGCACTGAAGTGTAC 2535
Db 2552 TTCTTCAATATGATGCAGCTTAAGGTGAACTTCAACCACAACATCTGCACTGAAGTATGT 2611
Qy 2536 CATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATCGTCTCCACGCTCTTC 2595
Db 2612 CATAAAAGTATATCCAGACGTTGCACGCGTCCAGTCACGGCCATCGTGTCTACGTTGCAC 2671
Qy 2596 TACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATCATCATAGATACCACA 2655
Db 2672 TACGGAGGCAAGATGCGCACGACCAACCCGTGCAACAAACCCATAATCATAGACACCACA 2731
Qy 2656 GGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGAGGATGGGTGAAA 2715





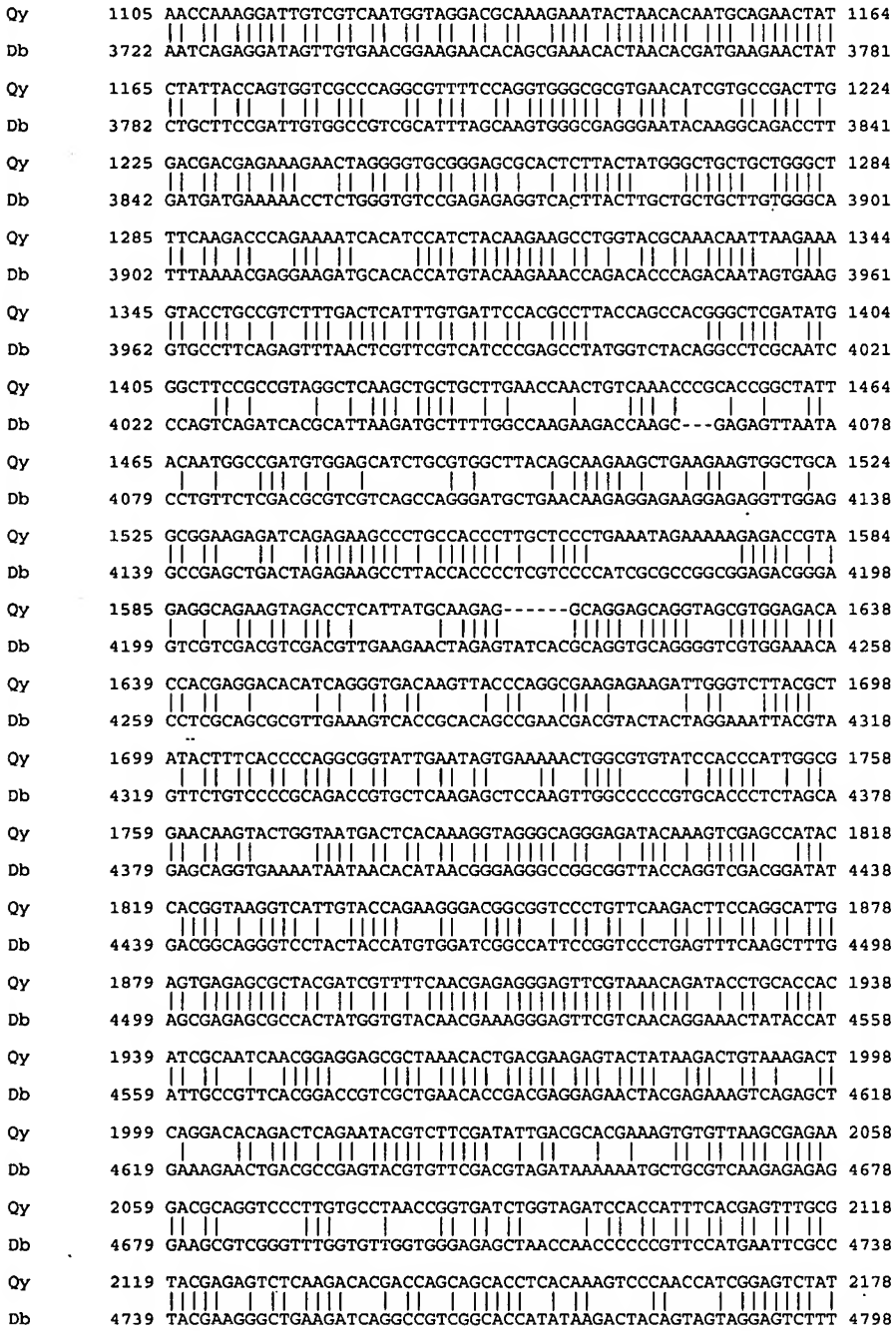
```
Qy       4861 TGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGC 4920
             |||||||| || |   | |||||| |||     |     || ||||||||| || |||||
Db       4949 TGCTCATCTTTTCCCCTCCCGAAATACCATGTAGATGGGGTGCAGAAGGTAAAGTGCGAG 5008

Qy       4921 AAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGCGGAA 4980
              ||      |||||||           |||||  | |  ||    ||  ||||||| || | |
Db       5009 AAGGTTCTCCTGTTCGACCCGACGGTACCTTCAGTGGTTAGTCCGCGGAAGTATGCCGCA 5068
```

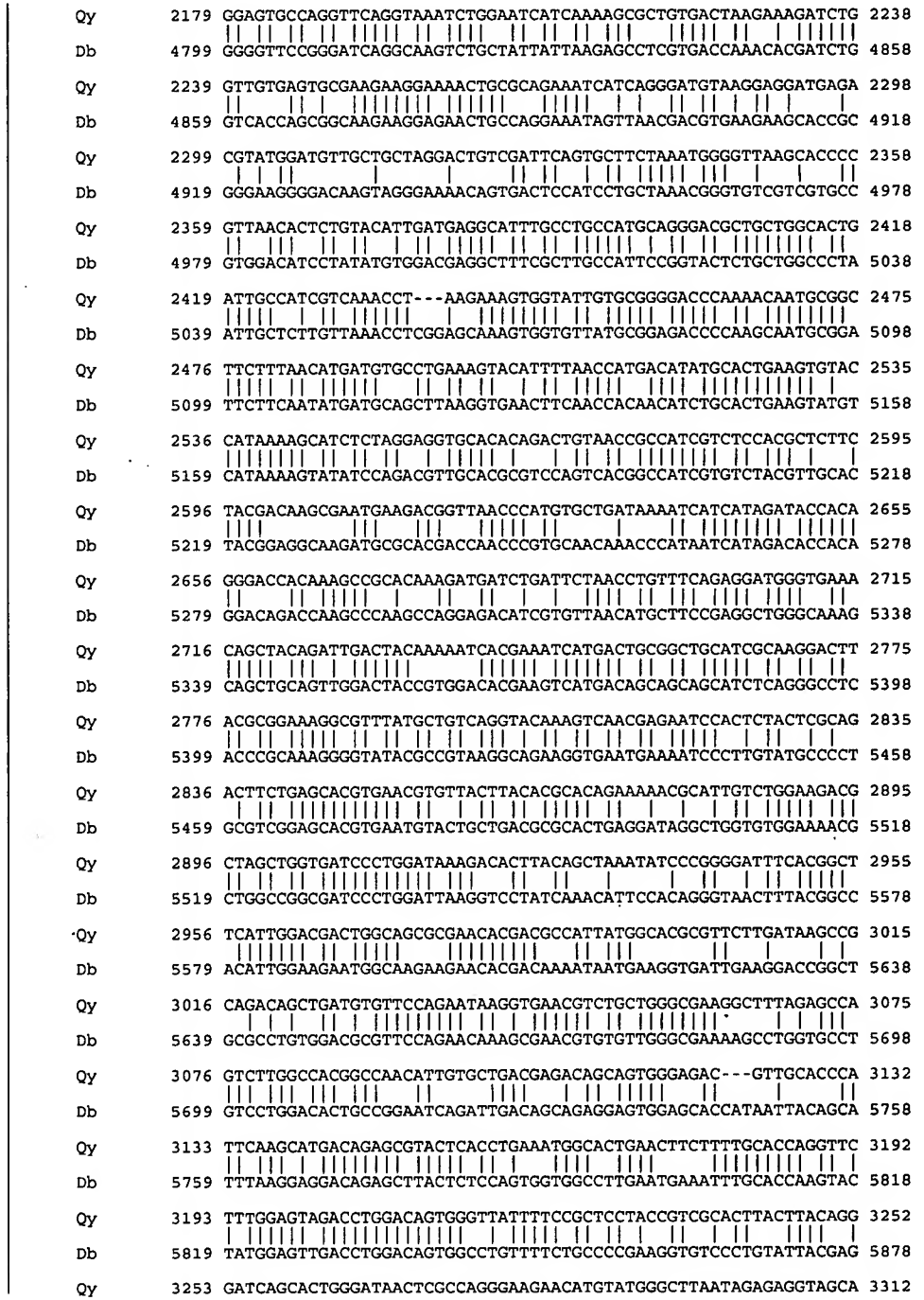
RESULT 11
AAX76582
ID AAX76582 standard; DNA; 15538 BP.
XX
AC AAX76582;
XX
DT 11-AUG-1999 (first entry)
XX
DE Plasmid pMP44 nucleotide sequence.
XX
KW Plasmid pMP44; hepatitis delta ribozyme; alphavirus vector; vaccine;
KW paramyxovirus; respiratory syncytial virus fusion protein; RSV F;
KW detection; infection; immune response; immunisation; ss.
XX
OS Synthetic.
XX
PN WO9925858-A1.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-CA01064.
XX
PR 14-NOV-1997; 97US-0065791.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Li X, Parrington M;
XX
DR WPI; 1999-370764/31.
XX
PT Improved alphavirus vector containing DNA encoding a paramyxovirus
PT protein
XX
PS Claim 19; Fig 3; 61pp; English.
XX
CC The present invention describes an improved alphavirus vector containing
CC DNA encoding an immunogenic paramyxovirus protein and DNA to enhance the
CC immunoprotective ability of the paramyxovirus protein in vivo. The vector
CC comprises: (1) a first DNA sequence which is complementary to at least
CC part of an alphavirus RNA genome and including the complement of the
CC complete alphavirus RNA genome replication regions to permit in vivo
CC replication; and (2) a second DNA sequence encoding a paramyxovirus
CC protein or protein fragment that generates antibodies that specifically
CC react with the paramyxovirus protein; where the second DNA sequence is
CC inserted into a region of the first DNA sequence which is non-essential
CC for replication, the first and second DNA sequences being under
CC transcriptional control of a promoter. The vector is used to immunize a
CC host against disease caused by infection with paramyxovirus, especially
CC respiratory syncytial virus. Human RSV is a major pathogen responsible
CC for severe respiratory tract infections in infants, young children and
CC the institutionalized elderly. The vector provides a protective immune
CC response using a lower dose and less time than vectors in US8476397 and
CC US8896500. Additionally the vectors using native RSV F produce
CC protective immune responses in the absence of pretreatment of the animal
CC model with cardiotoxin, a material known to increase the uptake of DNA
CC and enhance the immune response. The present sequence represents a
CC specifically claimed plasmid pMP44 nucleotide sequence from the present
CC invention.
XX
SQ Sequence 15538 BP; 4313 A; 3774 C; 3837 G; 3614 T; 0 other;

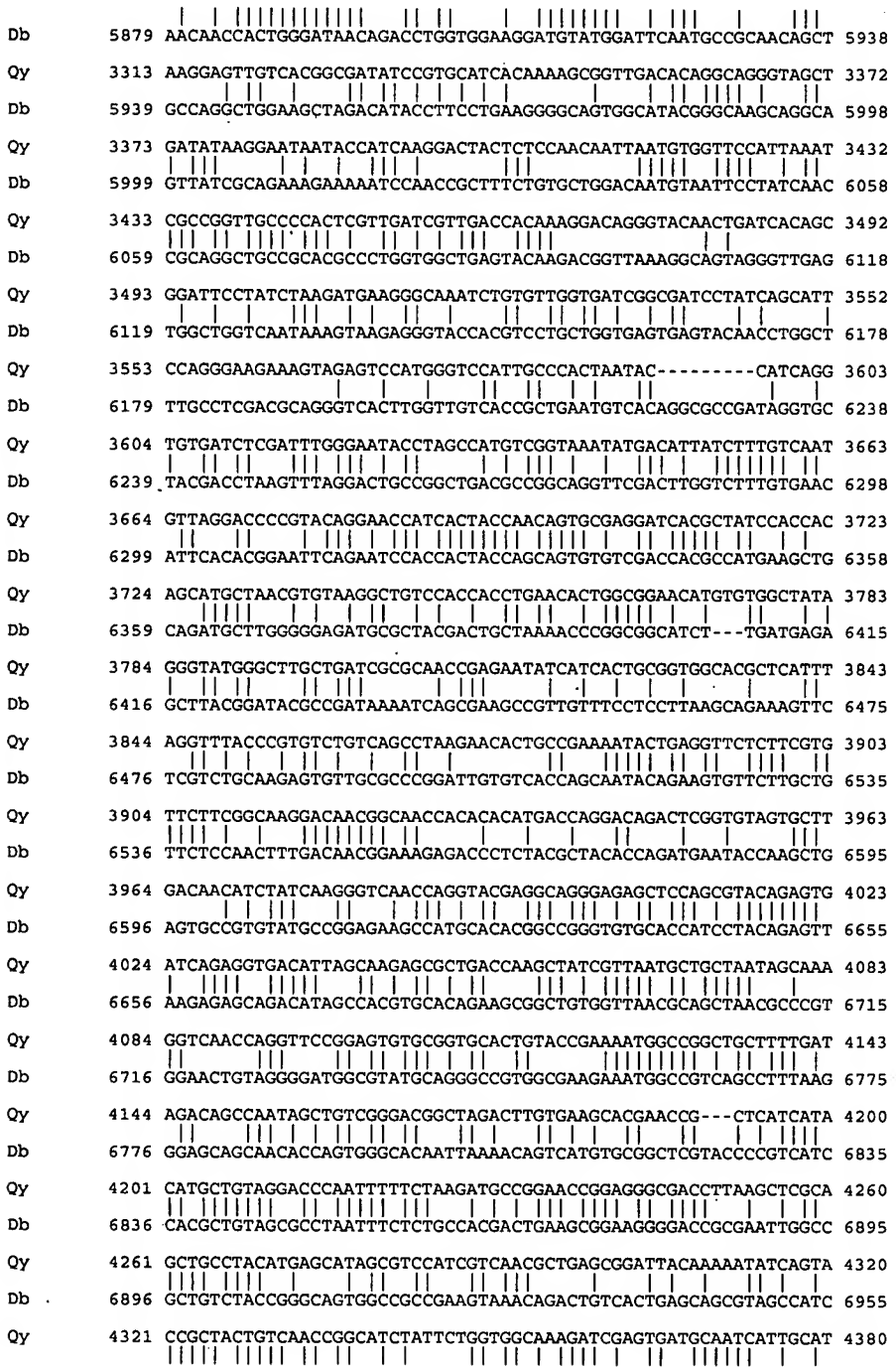
Query Match 31.3%; Score 1564; DB 20; Length 15538;
Best Local Similarity 58.3%; Pred. No. 0;

Matches 2904; Conservative 0; Mismatches 2040; Indels 36; Gaps 8;

```
Qy         31 AGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGTCGTTACAGCGGACGTTT 90
              | | | || |||||  | || ||||||||||| |   ||||||| || |||  | | |||
Db       2642 AAAGTGCATGTTGATATTGAGGCTGACAGCCCATTCATCAAGTCTTTGCAGAAGGCATTT 2701

Qy         91 CCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACCATGCCAATGCCAGAGCG 150
              ||     || ||| | ||  ||  |||||||||     |||||||||| ||||||||||
Db       2702 CCGTCGTTCGAGGTGGAGTCATTGCAGGTCACACCAAATGACCATGCAAATGCCAGAGCA 2761

Qy        151 TTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACCGGGACCAAGTTATCTTG 210
              ||||||||  |||| || ||  | || |||   ||     |||   |||  | | ||||||
Db       2762 TTTTCGCACCTGGCTACCAAATTGATCGAGCAGGAGACTGACAAAGACACACTCATCTTG 2821

Qy        211 GACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACCGCTATCATTGTATCTGC 270
              || || || ||||||||   |||          || |  |||   || || ||  | |||
Db       2822 GATATCGGCAGTGCGCCTTCCAGGAGAATGATGTCTACGCACAAATACCACTGCGTATGC 2881

Qy        271 CCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGGTATGCAGAAAGACT------T 324
              ||||||   ||||| |||||||| || || ||  |  | || ||| | | |||      |
Db       2882 CCTATGCGCAGCGCAGAAGACCCCGAAAGGCTCGATAGCTACGCAAAGAAACTGGCAGCG 2941

Qy        325 AAGAAAAGTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACCTGCTGGAAGTC 384
                  |  |  |   || |  | || ||   ||    | |||||| |  |||
Db       2942 GCCTCCGGGAAGGTGCTGGATAGAGAGATCGCAGGAAAAATCACCGACCTGCAGACCGTC 3001

Qy        385 ATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACGCCACGTGTAGG 444
              ||| | || |||||||| ||  ||||  |  | ||  |||| |||||| |||||| |
Db       3002 ATGGCTACGCCAGACGCTGAATCTCCTACCTTTGCCTGCATACAGACGTCACGTGTCGT 3061

Qy        445 TACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCACCGACATCAATC 504
                  |  || |||||||| || |||| || || |||||||| ||||  |
Db       3062 ACGGCAGCCGAAGTGGCCGTATACCAGGACGTGTATGCTGTACATGCACCAACATCGCTG 3121

Qy        505 TACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTGACACGACCCCT 564
              ||||| |||||| | ||||| || || ||    || |||| || |||||||| ||||
Db       3122 TACCATCAGGCGATGAAAGGTGTCAGAACGGCGTATTGGATTGGGTTTGACACCACCCCG 3181

Qy        565 TTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACTGGGCTGACGAG 624
              |||||||   |     | |||||  | || || || |||  ||| |||||||| ||||||
Db       3182 TTTATGTTTGACGCGCTAGCAGGCGCGTATCCAACCTACGCCACAAACTGGGCCGACGAG 3241

Qy        625 AGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGGAGAGCAGGCTT 684
              || ||  | ||  | |||||| || ||  ||     ||     |   ||| | || ||
Db       3242 CAGGTGTTACAGGCCAGGAACATAGGACTGTGTGCAGCATCCTTGACTGAGGGAAGACTC 3301

Qy        685 GGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGATCATATTCTCG 744
              || ||||| || || ||  | ||||||   |  |||||   | |   |||| ||||||
Db       3302 GGCAAACTGTCCATTCTCCGCAAGAAGCAATTGAAACCTTGCGACACAGTCATGTTCTCG 3361

Qy        745 GTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGCATCTTCCAAAC 804
              || || || ||| | ||||| ||    |||   ||  |  | ||||||||  | ||   |
Db       3362 GTAGGATCTACATTGTACACTGAGAGCAGAAAGCTACTGAGGAGCTGGCACTTACCCTCC 3421

Qy        805 GTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTGTCAGCTGT 864
              || |||||| |||||||  ||       ||| ||  |||| || |  ||||| ||    |||
Db       3422 GTATTCCACCTGAAAGGTAAACAATCCTTTACCTGTAGGTGCGATACCATCGTATCATGT 3481

Qy        865 GAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAGTTGAGAAC 924
              ||||||||||  | || || || || ||  |||| || || |||||||||  |
Db       3482 GAAGGGTACGTAGTTAAGAAAATCACTATGTGCCCCGGCCTGTACGGTAAAACGGTAGGG 3541

Qy        925 TTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATACGCTGCGC 984
              |  ||    ||    ||    ||||| ||| |    |||||  ||||||  || |
Db       3542 TACGCCGTGACGTATCACGCGGAGGGATTCCTAGTGTGCAAGACCACAGACACTGTCAAA 3601

Qy        985 GGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCGATCAGATG 1044
              || || || || || ||  |||| || || || || || ||  | || | || ||||| |||
Db       3602 GGAGAAAGAGTCTCATTCCCTGTATGCACCTACGTCCCCTCAACCATCTGTGATCAAATG 3661

Qy       1045 ACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTATTGGTTGGGCTC 1104
              || || || || || || ||||| |    ||| |||||||| ||  | || || ||  |
Db       3662 ACTGGCATACTAGCGACCGACGTCACACCGGAGGACGCACAGAAGTTGTTAGTGGGATTG 3721
```



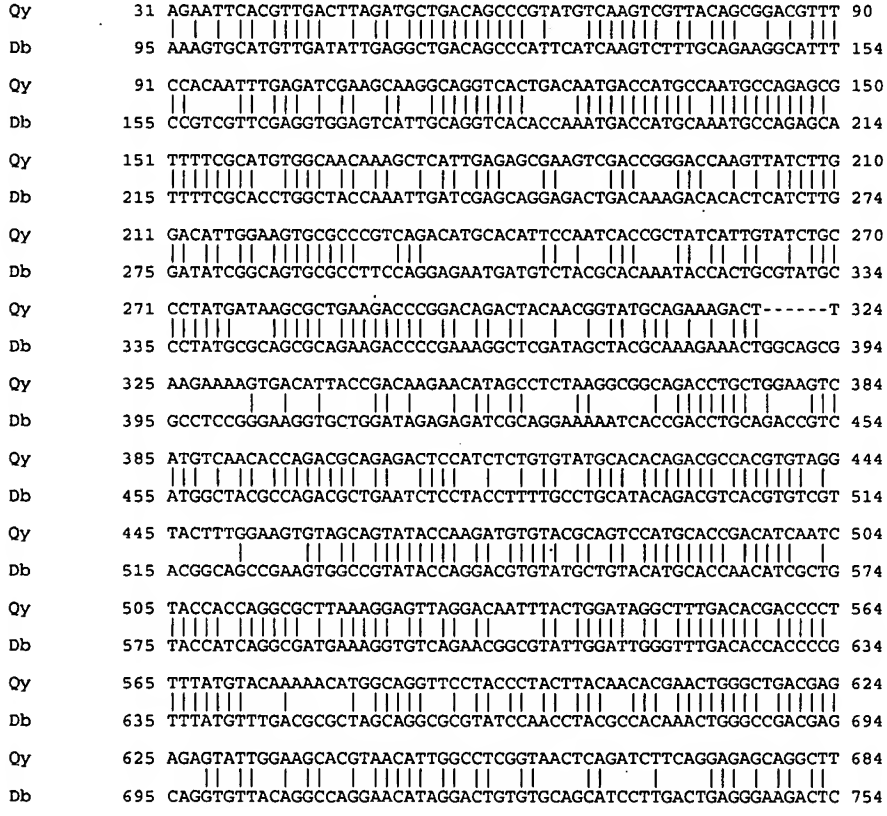


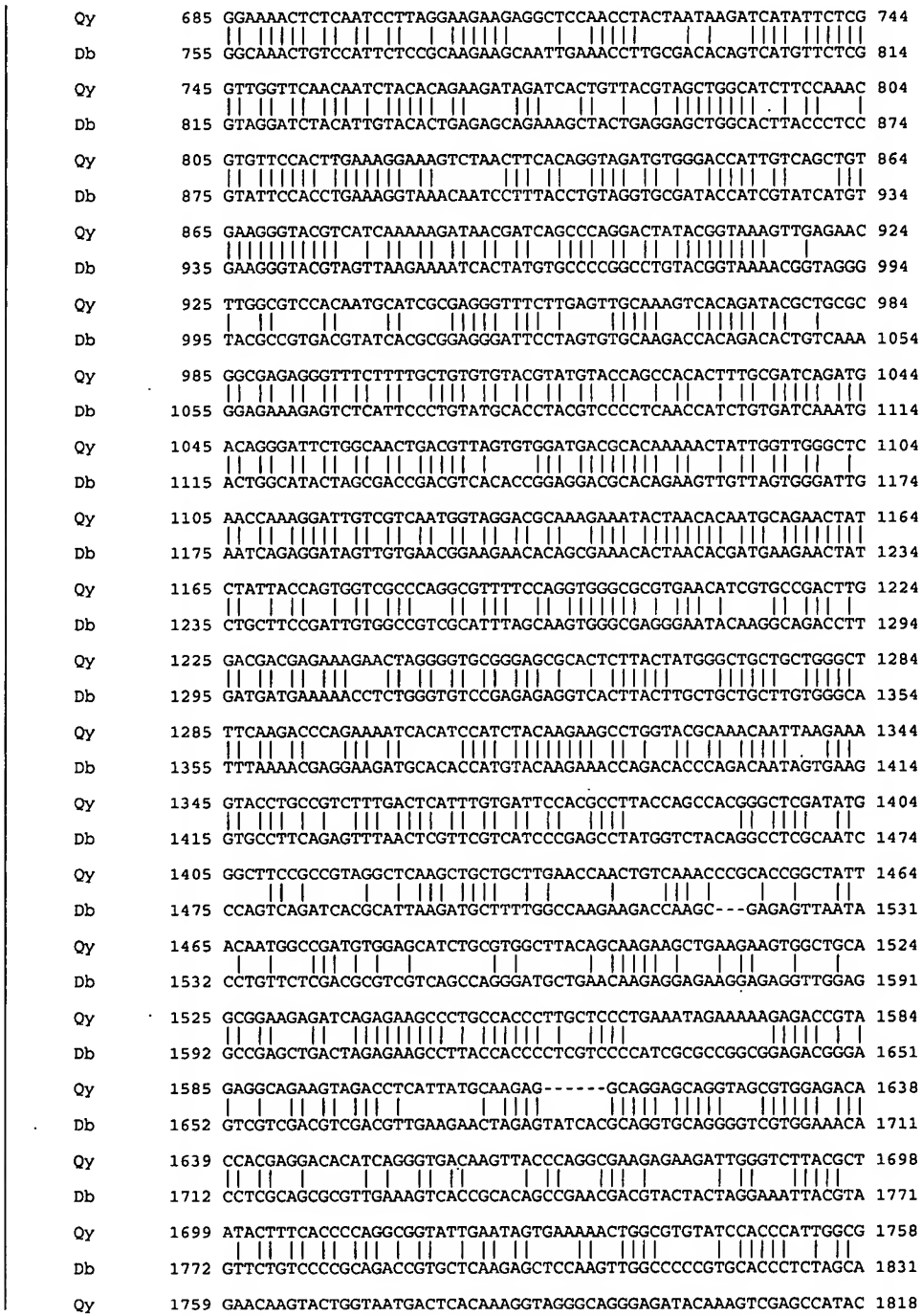


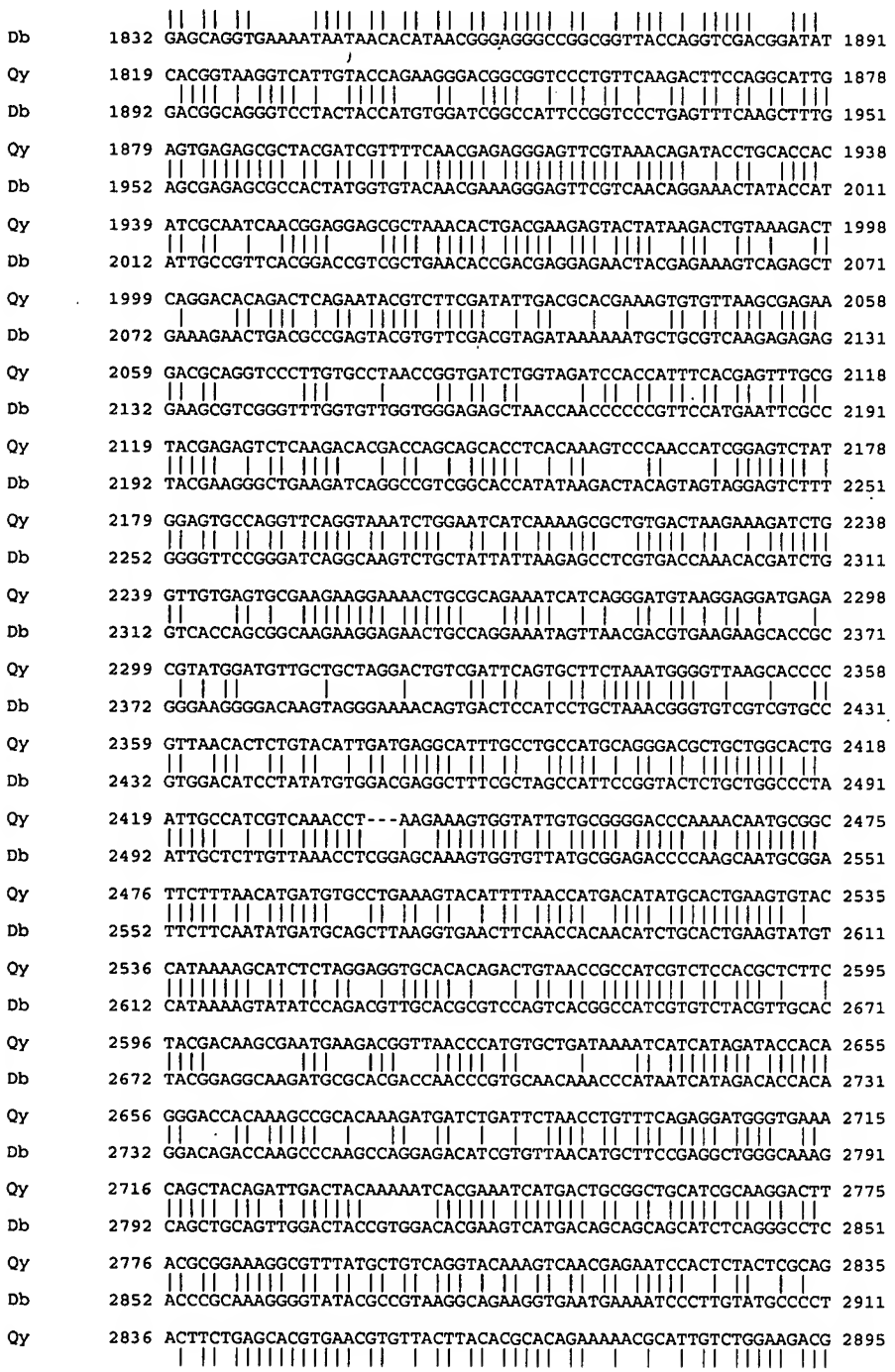
RESULT 12
ABN86687
ID ABN86687 standard; DNA; 11489 BP.
XX
AC ABN86687;
XX
DT 05-NOV-2002 (first entry)
XX
DE Nucleotide sequence of a pSCA1 suicide DNA vector.
XX
KW Major histocompatibility complex; MHC; antigen presenting cell; APC;
KW antigen; cytostatic; virucide; gene therapy; CD8; vaccine; therapeutic;
KW cancer; viral infection; ds.
XX
OS Synthetic.
XX
PN WO200261113-A2.
XX
PD 08-AUG-2002.
XX
PF 01-FEB-2002; 2002WO-US02598.
XX
PR 01-FEB-2001; 2001US-265334P.
XX
PA (UYJO) UNIV JOHNS HOPKINS.
XX
PI Wu T, Hung C;
XX
DR WPI; 2002-619261/66.
XX

PT Nucleic acid molecule encoding a fusion polypeptide that promotes
PT processing via the Major Histocompatibility Complex class I pathway
PT and/or promotes activity of an antigen presenting cell, useful as
PT vaccine for cancer and viral infections -
XX
PS Claim 32; Page 25-27; 127pp; English.
XX
CC The invention relates to a new nucleic acid molecule (I) encoding a
CC fusion polypeptide useful as a vaccine composition. (I) comprises a first
CC nucleic acid sequence encoding a first polypeptide or peptide that
CC promotes processing via the Major Histocompatibility Complex (MHC) class
CC I pathway (MHC-I-PP) and/or promotes development or activity of an
CC antigen presenting cell (APC). The second nucleic acid sequence is linked
CC in frame to the first nucleic acid sequence or to a linker nucleic acid
CC sequence and encodes an antigenic polypeptide or peptide. The methods and
CC compositions of the present invention are useful as therapeutic vaccine
CC for cancer and for major viral infections, such as hepatoma and cervical
CC cancer, that cause morbidity and mortality. They can also be used in
CC treating animal diseases, such as equine herpesvirus, bovine viruses,
CC Marek's disease, retroviral and lentiviral diseases and rabies, in the
CC veterinary medicine context. The present sequence represents the
CC nucleotide sequence of a pSCA1 suicide DNA vector.
XX
SQ Sequence 11489 BP; 3105 A; 2908 C; 2988 G; 2488 T; 0 other;

Query Match 31.2%; Score 1562.4; DB 24; Length 11489;
Best Local Similarity 58.3%; Pred. No. 0;
Matches 2903; Conservative 0; Mismatches 2041; Indels 36; Gaps 8;





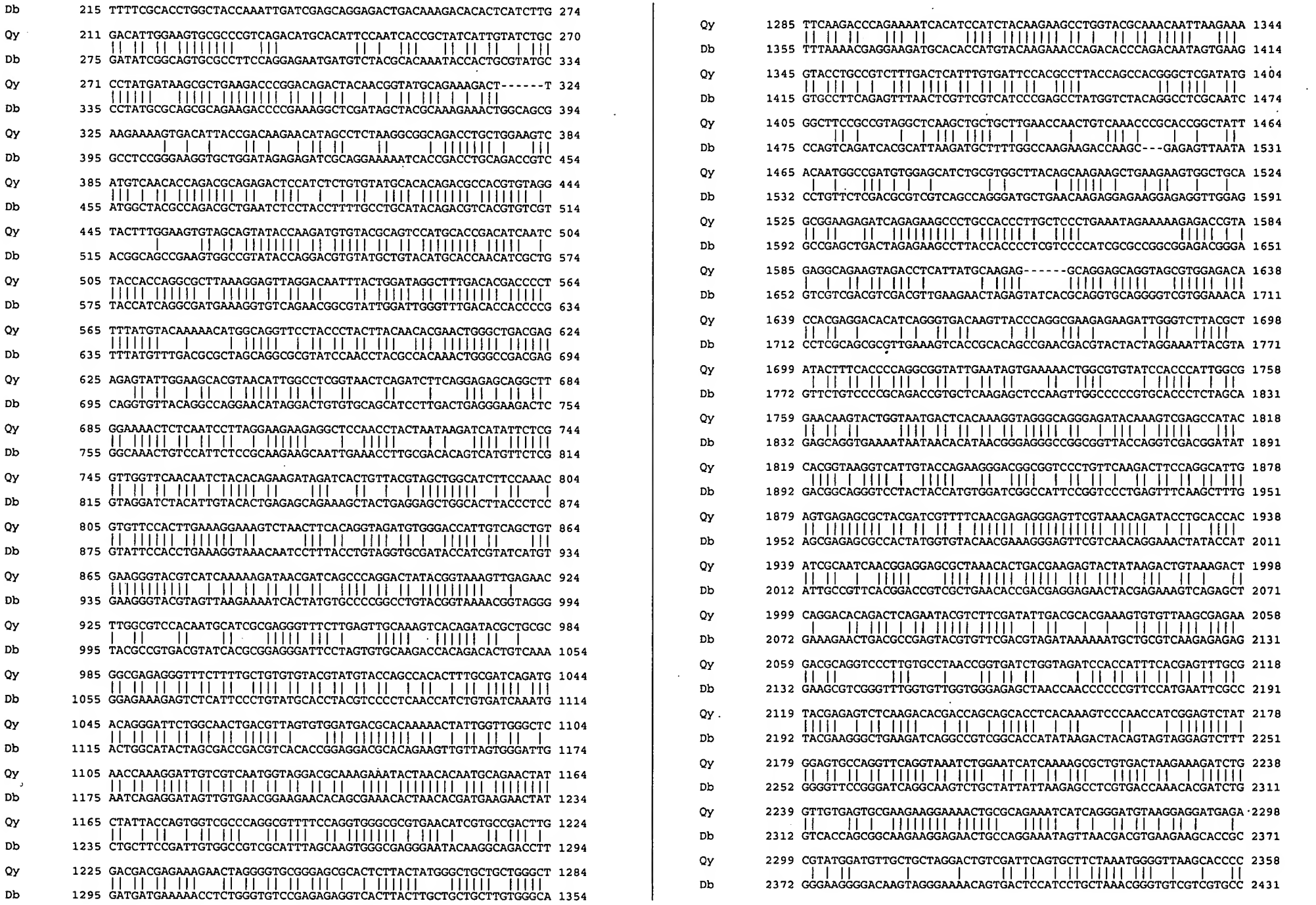

Db 1832 GAGCAGGTGAAAATAATAACACATAACGGGAGGGCCGGCGGTTACCAGGTCGACGGATAT 1891
Qy 1819 CACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAAGACTTCCAGGCATTG 1878
Db 1892 GACGGCAGGGTCCTACTACCATGTGGATCGGCCATTCCGGTCCCTGAGTTTCAAGCTTTG 1951
Qy 1879 AGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAACAGATACCTGCACCAC 1938
Db 1952 AGCGAGAGCGCCACTATGGTGTACAACGAAAGGGAGTTCGTCAACAGGAAACTATACCAT 2011
Qy 1939 ATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTATAAGACTGTAAAGACT 1998
Db 2012 ATTGCCGTTCACGGACCGTCGCTGAACACCGACGAGGAGAACTACGAGAAAGTCAGAGCT 2071
Qy 1999 CAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAGTGTGTTAAGCGAGAA 2058
Db 2072 GAAAGAACTGACGCCGAGTACGTGTTCGACGTAGATAAAAAATGCTGCGTCAAGAGAGAG 2131
Qy 2059 GACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCATTTCACGAGTTTGCG 2118
Db 2132 GAAGCGTCGGGTTTGGTGTTGGTGGGAGAGCTAACCAACCCCCCGTTCCATGAATTCGCC 2191
Qy 2119 TACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCAACCATCGGAGTCTAT 2178
Db 2192 TACGAAGGGCTGAAGATCAGGCCGTCGGCACCATATAAGACTACAGTAGTAGGAGTCTTT 2251
Qy 2179 GGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTGACTAAGAAAGATCTG 2238
Db 2252 GGGGTTCCGGGATCAGGCAAGTCTGCTATTATTAAGAGCCTCGTGACCAAACACGATCTG 2311
Qy 2239 GTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGATGTAAGGAGGATGAGA 2298
Db 2312 GTCACCAGCGGCAAGAAGGAGAACTGCCAGGAAATAGTTAACGACGTGAAGAAGCACCGC 2371
Qy 2299 CGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAATGGGGTTAAGCACCCC 2358
Db 2372 GGGAAGGGGACAAGTAGGGAAAACAGTGACTCCATCCTGCTAAACGGGTGTCGTCGTGCC 2431
Qy 2359 GTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGGACGCTGCTGGCACTG 2418
Db 2432 GTGGACATCCTATATGTGGACGAGGCTTTCGCTAGCCATTCCGGTACTCTGCTGGCCCTA 2491
Qy 2419 ATTGCCATCGTCAAACCT---AAGAAAGTGGTATTGTGCGGGGACCCAAAACAATGCGGC 2475
Db 2492 ATTGCTCTTGTTAAACCTCGGAGCAAAGTGGTGTTATGCGGAGACCCCAAGCAATGCGGA 2551
Qy 2476 TTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGCACTGAAGTGTAC 2535
Db 2552 TTCTTCAATATGATGCAGCTTAAGGTGAACTTCAACCACAACATCTGCACTGAAGTATGT 2611
Qy 2536 CATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATCGTCTCCACGCTCTTC 2595
Db 2612 CATAAAAGTATATCCAGACGTTGCACGCGTCCAGTCACGGCCATCGTGTCTACGTTGCAC 2671
Qy 2596 TACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATCATCATAGATACCACA 2655
Db 2672 TACGGAGGCAAGATGCGCACGACCAACCCGTGCAACAAACCCATAATCATAGACACCACA 2731
Qy 2656 GGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGAGGATGGGTGAAA 2715
Db 2732 GGACAGACCAAGCCCAAGCCAGGAGACATCGTGTTAACATGCTTCCGAGGCTGGGCAAAG 2791
Qy 2716 CAGCTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCTGCATCGCAAGGACTT 2775
Db 2792 CAGCTGCAGTTGGACTACCGTGGACACGAAGTCATGACAGCAGCAGCATCTCAGGGCCTC 2851
Qy 2776 ACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCACTCTACTCGCAG 2835
Db 2852 ACCCGCAAAGGGGTATACGCCGTAAGGCAGAAGGTGAATGAAAATCCCTTGTATGCCCCT 2911
Qy 2836 ACTTCTGAGCACGTGAACGTGTTACTTACACGCACAGAAAAACGCATTGTCTGGAAGACG 2895
Db 2912 GCGTCGGAGCACGTGAATGTACTGCTGACGCGCACTGAGGATAGGCTGGTGTGGAAAACG 2971
Qy 2896 CTAGCTGGTGATCCCTGGATAAAGACACTTACAGCTAAATATCCCGGGGATTTCACGGCT 2955
Db 2972 CTGGCCGGCGATCCCTGGATTAAGGTCCTATCAAACATTCCACAGGGTAACTTTACGGCC 3031
Qy 2956 TCATTGGACGACTGGCAGCGCGAACACGACGCCATTATGGCACGCGTTCTTGATAAGCCG 3015
Db 3032 ACATTGGAAGAATGGCAAGAAGAACACGACAAAATAATGAAGGTGATTGAAGGACCGGCT 3091
Qy 3016 CAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAGGCTTTAGAGCCA 3075
Db 3092 GCGCCTGTGGACGCGTTCCAGAACAAAGCGAACGTGTGTTGGGCGAAAAGCCTGGTGCCT 3151
Qy 3076 GTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAGAC---GTTGCACCCA 3132
Db 3152 GTCCTGGACACTGCCGGAATCAGATTGACAGCAGAGGAGTGGAGCACCATAATTACAGCA 3211
Qy 3133 TTCAAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTTTGCACCAGGTTC 3192
Db 3212 TTTAAGGAGGACAGAGCTTACTCTCCAGTGGTGGCCTTGAATGAAATTTGCACCAAGTAC 3271
Qy 3193 TTTGGAGTAGACCTGGACAGTGGGTTATTTTCCGCTCCTACCGTCGCACTTACTTACAGG 3252
Db 3272 TATGGAGTTGACCTGGACAGTGGCCTGTTTTCTGCCCCGAAGGTGTCCCTGTATTACGAG 3331
Qy 3253 GATCAGCACTGGGATAACTCGCCAGGGAAGAACATGTATGGGCTTAATAGAGAGGTAGCA 3312
Db 3332 AACAACCACTGGGATAACAGACCTGGTGGAAGGATGTATGGATTCAATGCCGCAACAGCT 3391
Qy 3313 AAGGAGTTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACAGGCAGGGTAGCT 3372
Db 3392 GCCAGGCTGGAAGCTAGACATACCTTCCTGAAGGGGCAGTGGCATACGGGCAAGCAGGCA 3451
Qy 3373 GATATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTGGTTCCATTAAAT 3432
Db 3452 GTTATCGCAGAAAGAAAAATCCAACCGCTTTCTGTGCTGGACAATGTAATTCCTATCAAC 3511
Qy 3433 CGCCGGTTGCCCCACTCGTTGATCGTTGACCACAAAGGACAGGGTACAACTGATCACAGC 3492
Db 3512 CGCAGGCTGCCGCACGCCCTGGTGGCTGAGTACAAGACGGTTAAAGGCAGTAGGGTTGAG 3571
Qy 3493 GGATTCCTATCTAAGATGAAGGGCAAATCTGTGTTGGTGATCGGCGATCCTATCAGCATT 3552
Db 3572 TGGCTGGTCAATAAAGTAAGAGGGTACCACGTCCTGCTGGTGAGTGAGTACAACCTGGCT 3631
Qy 3553 CCAGGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATAC---------CATCAGG 3603
Db 3632 TTGCCTCGACGCAGGGTCACTTGGTTGTCACCGCTGAATGTCACAGGCGCCGATAGGTGC 3691
Qy 3604 TGTGATCTCGATTTGGGAATACCTAGCCATGTCGGTAAATATGACATTATCTTTGTCAAT 3663
Db 3692 TACGACCTAAGTTTAGGACTGCCGGCTGACGCCGGCAGGTTCGACTTGGTCTTTGTGAAC 3751
Qy 3664 GTTAGGACCCCGTACAGGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCACCAC 3723
Db 3752 ATTCACACGGAATTCAGAATCCACCACTACCAGCAGTGTGTCGACCACGCCATGAAGCTG 3811
Qy 3724 AGCATGCTAACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCTATA 3783
Db 3812 CAGATGCTTGGGGGAGATGCGCTACGACTGCTAAAACCCGGCGGCATCT---TGATGAGA 3868
Qy 3784 GGGTATGGGCTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCACGCTCATTT 3843
Db 3869 GCTTACGGATACGCCGATAAAATCAGCGAAGCCGTTGTTTCCTCCTTAAGCAGAAAGTTC 3928
Qy 3844 AGGTTTACCCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTCGTG 3903
Db 3929 TCGTCTGCAAGAGTGTTGCGCCCGGATTGTGTCACCAGCAATACAGAAGTGTTCTTGCTG 3988
Qy 3904 TTCTTCGGCAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTGCTT 3963
Db 3989 TTCTCCAACTTTGACAACGGAAAGAGACCCTCTACGCTACACCAGATGAATACCAAGCTG 4048

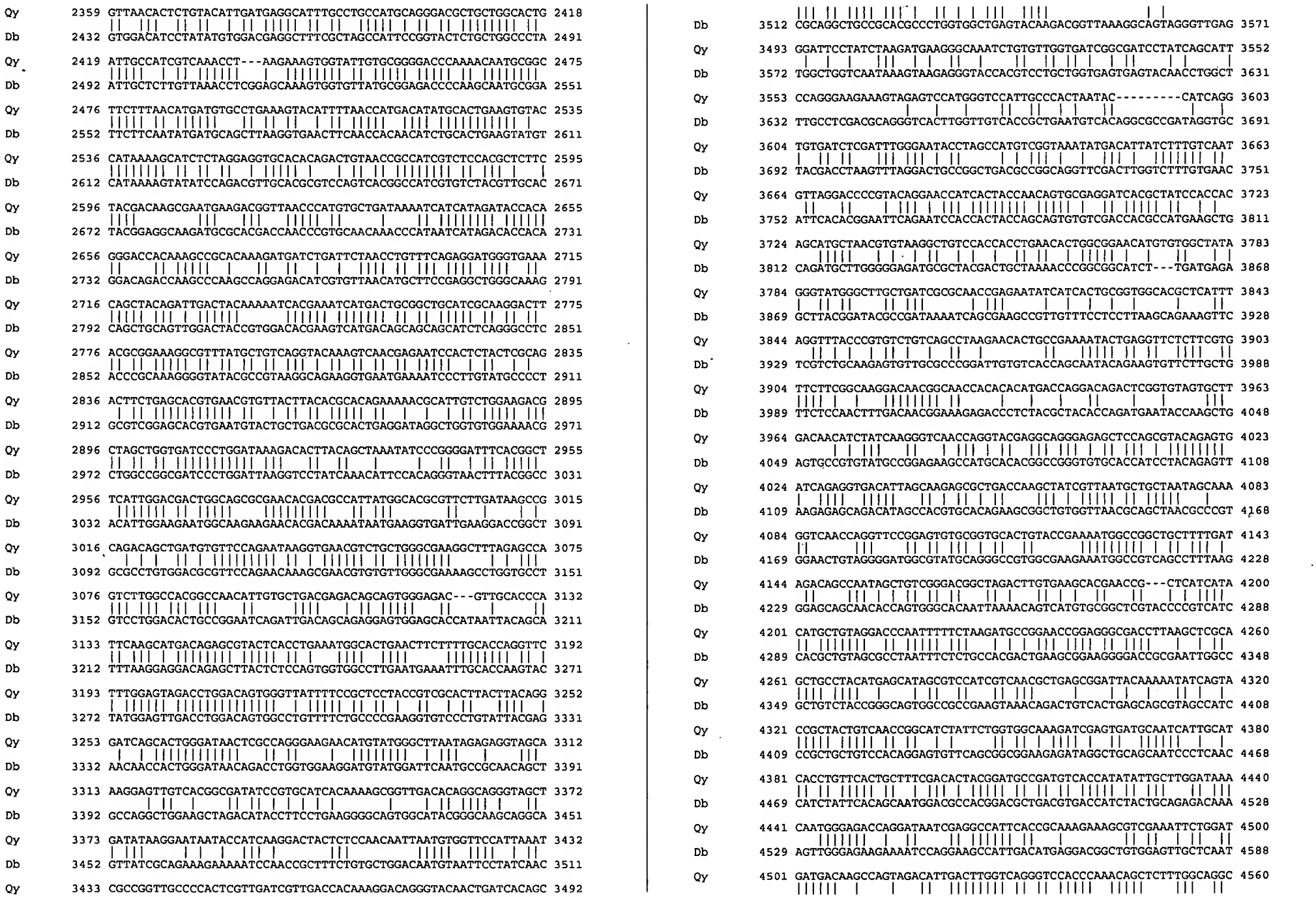
```
Qy       3964 GACAACATCTATCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGTACAGAGTG 4023
Db       4049 AGTGCCGTGTATGCCGGAGAAGCCATGCACACGGCCGGGTGTGCACCATCCTACAGAGTT 4108

Qy       4024 ATCAGAGGTGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGCAAA 4083
Db       4109 AAGAGAGCAGACATAGCCACGTGCACAGAAGCGGCTGTGGTTAACGCAGCTAACGCCCGT 4168

Qy       4084 GGTCAACCAGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTTGAT 4143
Db       4169 GGAACTGTAGGGGATGGCGTATGCAGGGCCGTGGCGAAGAAATGGCCGTCAGCCTTTAAG 4228

Qy       4144 AGACAGCCAATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACCG---CTCATCATA 4200
Db       4229 GGAGCAGCAACACCAGTGGGCACAATTAAAACAGTCATGTGCGGCTCGTACCCCGTCATC 4288

Qy       4201 CATGCTGTAGGACCCAATTTTTCTAAGATGCCCGGAACCGGAGGGCGACCTTAAGCTCGCA 4260
Db       4289 CACGCTGTAGCGCCTAATTTCTCTGCCACGACTGAAGCGGAAGGGGACCGCGAATTGGCC 4348

Qy       4261 GCTGCCTACATGAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTA 4320
Db       4349 GCTGTCTACCGGGCAGTGGCCGCCGAAGTAAACAGACTGTCACTGAGCAGCGTAGCCATC 4408

Qy       4321 CCGCTACTGTCAACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCAT 4380
Db       4409 CCGCTGCTGTCCACAGGAGTGTTCAGCGGCGGAAGAGATAGGCTGCAGCAATCCCTCAAC 4468

Qy       4381 CACCTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAA 4440
Db       4469 CATCTATTCACAGCAATGGACGCCACGGACGCTGACGTGACCATCTACTGCAGAGACAAA 4528

Qy       4441 CAATGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATTCTGGAT 4500
Db       4529 AGTTGGGAGAAGAAAATCCAGGAAGCCATTGACATGAGGACGGCTGTGGAGTTGCTCAAT 4588

Qy       4501 GATGACAAGCCAGTAGACATTGACTTGGTCAGGGTCCACCCAAACAGCTCTTTGGCAGGC 4560
Db       4589 GATGACGTGGAGCTGACCACAGACTTGGTGAGAGTGCACCCGGACAGCAGCCTGGTGGGT 4648

Qy       4561 AGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTGGAAGGTACACGATTC 4620
Db       4649 CGTAAGGGCTACAGTACCACTGACGGGTCGCTGTACTCGTACTTTGAAGGTACGAAATTC 4708

Qy       4621 CATCAGACCGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCCAACAAATCTGAGGCT 4680
Db       4709 AACCAGGCTGCTATTGATATGGCAGAGATACTGACGTTGTGGCCCAGACTGCAAGAGGCA 4768

Qy       4681 AATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGC 4740
Db       4769 AACGAACAGATATGCCTATACGCGCTGGGCGAAACAATGGACAACATCAGATCCAAATGT 4828

Qy       4741 CCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTAC 4800
Db       4829 CCGGTGAACGATTCCGATTCATCAACACCTCCCAGGACAGTGCCCTGCCTGTGCCGCTAC 4888

Qy       4801 GCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTA 4860
Db       4889 GCAATGACAGCAGAACGGATCGCCCGCCTTAGGTCACACCAAGTTAAAAGCATGGTGGTT 4948

Qy       4861 TGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGC 4920
Db       4949 TGCTCATCTTTTCCCCTCCCGAAATACCATGTAGATGGGGTGCAGAAGGTAAAGTGCGAG 5008

Qy       4921 AAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGCGGAA 4980
Db       5009 AAGGTTCTCCTGTTCGACCCGACGGTACCTTCAGTGGTTAGTCCGCGGAAGTATGCCGCA 5068
```

RESULT 13
ABN86690
ID ABN86690 standard; DNA; 13599 BP.
XX
AC ABN86690;
XX
DT 05-NOV-2002 (first entry)
XX
DE Nucleotide sequence of vector pSCA1-E7-Hsp70.
XX
KW Major histocompatibility complex; MHC; antigen presenting cell; APC;
KW antigen; cytostatic; virucide; gene therapy; CD8; vaccine; therapeutic;
KW cancer; viral infection; HPV; E7; heat shock protein 70; Hsp70; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT misc_feature 7405..9509
FT /*tag= a
FT /note= "E7-Hsp70 fusion sequence"
XX
PN WO200261113-A2.
XX
PD 08-AUG-2002.
XX
PF 01-FEB-2002; 2002WO-US02598.
XX
PR 01-FEB-2001; 2001US-265334P.
XX
PA (UYJO) UNIV JOHNS HOPKINS.
XX
PI Wu T, Hung C;
XX
DR WPI; 2002-619261/66.
XX
PT Nucleic acid molecule encoding a fusion polypeptide that promotes
PT processing via the Major Histocompatibility Complex class I pathway
PT and/or promotes activity of an antigen presenting cell, useful as
PT vaccine for cancer and viral infections -
XX
PS Claim 37; Page 34-37; 127pp; English.
XX
CC The invention relates to a new nucleic acid molecule (I) encoding a
CC fusion polypeptide useful as a vaccine composition. (I) comprises a first
CC nucleic acid sequence encoding a first polypeptide or peptide that
CC promotes processing via the Major Histocompatibility Complex (MHC) class
CC I pathway (MHC-I-PP) and/or promotes development or activity of an
CC antigen presenting cell (APC). The second nucleic acid sequence is linked
CC in frame to the first nucleic acid sequence or to a linker nucleic acid
CC sequence and encodes an antigenic polypeptide or peptide. The methods and
CC compositions of the present invention are useful as therapeutic vaccine
CC for cancer and for major viral infections, such as hepatoma and cervical
CC cancer, that cause morbidity and mortality. They can also be used in
CC treating animal diseases, such as equine herpesvirus, bovine viruses,
CC Marek's disease, retroviral and lentiviral diseases and rabies, in the
CC veterinary medicine context. The present sequence represents the
CC nucleotide sequence of vector pSCA1-E7-Hsp70 comprising the human
CC papillomavirus (HPV) E7 antigenic protein and M. tuberculosis heat
CC shock protein 70 (Hsp70) fusion sequence.
XX
SQ Sequence 13599 BP; 3582 A; 3516 C; 3647 G; 2854 T; 0 other;

Query Match 31.2%; Score 1562.4; DB 24; Length 13599;
Best Local Similarity 58.3%; Pred. No. 0;
Matches 2903; Conservative 0; Mismatches 2041; Indels 36; Gaps 8;

```
Qy         31 AGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGTCGTTACAGCGGACGTTT 90
Db         95 AAAGTGCATGTTGATATTGAGGCTGACAGCCCATTCATCAAGTCTTTGCAGAAGGCATTT 154

Qy         91 CCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACCATGCCAATGCCAGAGCG 150
Db        155 CCGTCGTTCGAGGTGGAGTCATTGCAGGTCACACCAAATGACCATGCAAATGCCAGAGCA 214

Qy        151 TTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACCGGGACCAAGTTATCTTG 210
```





```
Db       4589 GATGACGTGGAGCTGACCACAGACTTGGTGAGAGTGCACCCGGACAGCAGCCTGGTGGGT 4648

Qy       4561 AGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTGGAAGGTACACGATTC 4620

Db       4649 CGTAAGGGCTACAGTACCACTGACGGGTCGCTGTACTCGTACTTTGAAGGTACGAAATTC 4708

Qy       4621 CATCAGACCGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCCAACAAATCTGAGGCT 4680

Db       4709 AACCAGGCTGCTATTGATATGGCAGAGATACTGACGTTGTGGCCCAGACTGCAAGAGGCA 4768

Qy       4681 AATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGC 4740

Db       4769 AACGAACAGATATGCCTATACGCGCTGGGCGAAACAATGGACAACATCAGATCCAAATGT 4828

Qy       4741 CCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTAC 4800

Db       4829 CCGGTGAACGATTCCGATTCATCAACACCTCCCAGGACAGTGCCCTGCCTGTGCCGCTAC 4888

Qy       4801 GCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTA 4860

Db       4889 GCAATGACAGCAGAACGGATCGCCCGCCTTAGGTCACACCAAGTTAAAAGCATGGTGGTT 4948

Qy       4861 TGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGC 4920

Db       4949 TGCTCATCTTTTCCCCTCCCGAAATACCATGTAGATGGGGTGCAGAAGGTAAAGTGCGAG 5008

Qy       4921 AAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGCGGAA 4980

Db       5009 AAGGTTCTCCTGTTCGACCCGACGGTACCTTCAGTGGTTAGTCCGCGGAAGTATGCCGCA 5068
```

RESULT 14
AAQ26021
ID AAQ26021 standard; DNA; 11517 BP.
XX
AC AAQ26021;
XX
DT 25-MAR-2003 (updated)
DT 05-JAN-1993 (first entry)
XX
DE pSP6-SFV4 RNA transcript as DNA.
XX
KW Semliki forest virus; SFV; SP6 expression vector; RNA polymerase; ss.
XX
OS Synthetic.
XX

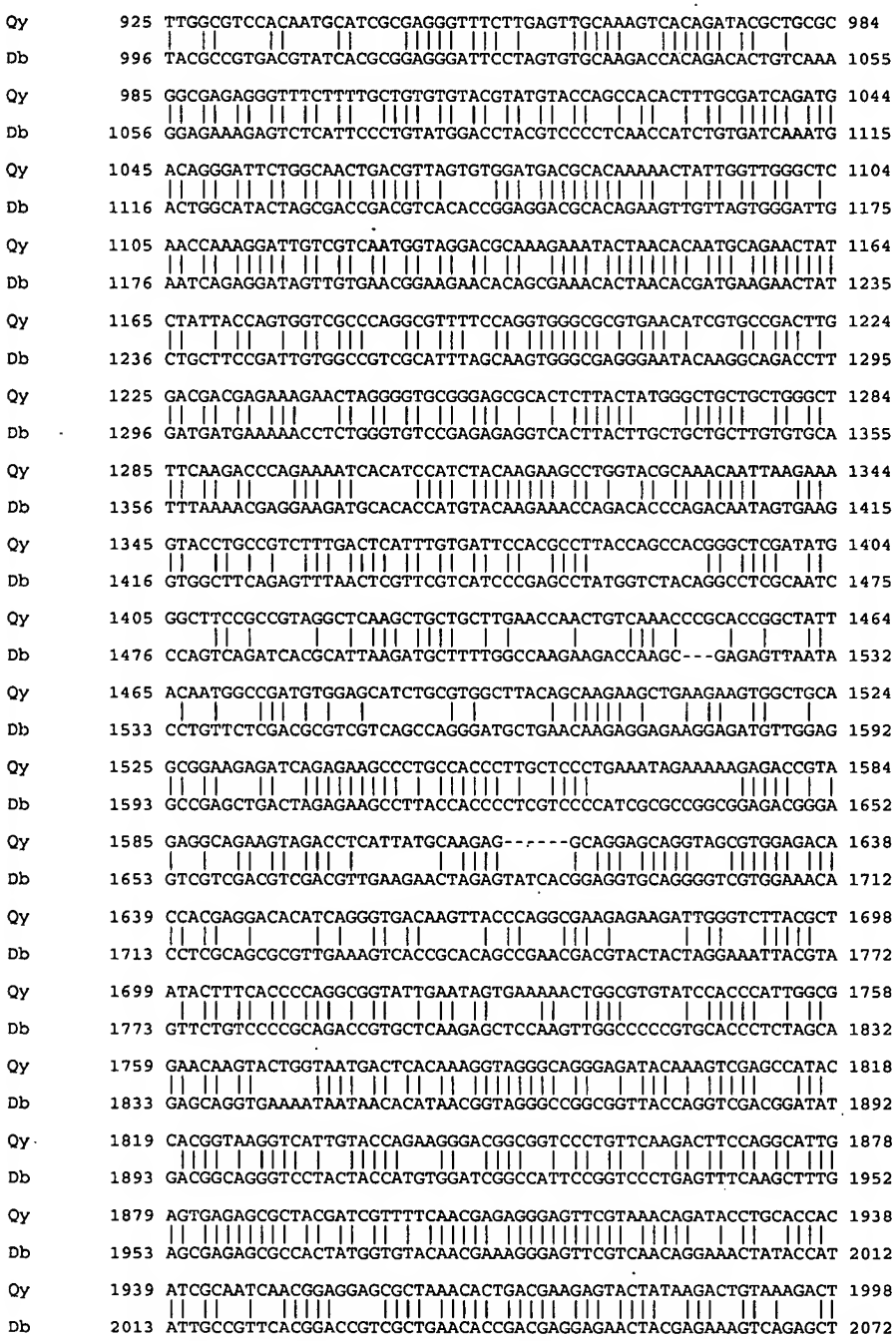
FH	Key	Location/Qualifiers
FT	CDS	87..7379
FT		/*tag= a
FT		/label= Non-structural_polyprotein
FT	CDS	7421..11182
FT		/*tag= b
FT		/label= Structural_polyprotein

XX
PN WO9210578-A1.
XX
PD 25-JUN-1992.
XX
PF 12-DEC-1991; 91WO-SE00855.
XX
PR 13-DEC-1990; 90SE-0003978.
XX
PA (BIOP-) BIOPTION AB.
XX
PI Garoff H, Liljestroem P;
XX
DR WPI; 1992-234633/28.
DR P-PSDB; AAR25138, AAR28337.
XX
PT RNA mol. derived from alphavirus RNA genome - chimeric alphavirus
PT antigen and vaccine for immunisation against viral infections
XX
PS Disclosure; Fig 5; 94pp; English.

XX
CC The sequence given contains a full length Semliki forest virus (SFV)
CC cDNA clone within an SP6 expression vector. The SP6 RNA polymerase
CC promoter allows in vitro transcription of full length and infectious
CC transcripts.
CC (Updated on 25-MAR-2003 to correct PN field.)
CC (Updated on 25-MAR-2003 to correct PI field.)
XX
SQ Sequence 11517 BP; 3130 A; 2985 C; 3091 G; 2311 T; 0 other;

Query Match 31.1%; Score 1554.4; DB 13; Length 11517;
Best Local Similarity 58.2%; Pred. No. 0;
Matches 2898; Conservative 0; Mismatches 2046; Indels 36; Gaps 8;

```
Qy         31 AGAATTCACGTTGACTTAGATGCTGACAGCCCGTATGTCAAGTCGTTACAGCGGACGTTT 90

Db         96 AAAGTGCATGTTGATATTGAGGCTGACAGCCCATTCATCAAGTCTTTGCAGAAGGCATTT 155

Qy         91 CCACAATTTGAGATCGAAGCAAGGCAGGTCACTGACAATGACCATGCCAATGCCAGAGCG 150

Db        156 CCGTCGTTCGAGGTGGAGTCATTGCAGGTCACACCAAATGACCATGCAAATGCCAGAGCA 215

Qy        151 TTTTCGCATGTGGCAACAAAGCTCATTGAGAGCGAAGTCGACCGGGACCAAGTTATCTTG 210

Db        216 TTTTCGCACCTGGCTACCAAATTGATCGAGCAGGAGACTGACAAAGACACACTCATCTTG 275

Qy        211 GACATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACCGCTATCATTGTATCTGC 270

Db        276 GATATCGGCAGTGCGCCTTCCAGGAGAATGATGTCTACGCACAAATACCACTGCGTATGC 335

Qy        271 CCTATGATAAGCGCTGAAGACCCGGACAGACTACAACGGTATGCAGAAAGACT------T 324

Db        336 CCTATGCGCAGCGCAGAAGACCCCGAAAGGCTCGATAGCTACGCAAAGAAACTGGCAGCG 395

Qy        325 AAGAAAAGTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACCTGCTGGAAGTC 384

Db        396 GCCTCCGGTAAGGTGCTGGATAGAGAGATCGCAGGAAAAATCACCGACCTGCAGACCGTC 455

Qy        385 ATGTCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACGCCACGTGTAGG 444

Db        456 ATGGCTACGCCAGACGCTGAATCTCCTACCTTTTGCCTGCATACAGACGTCACGTGTCGT 515

Qy        445 TACTTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCACCGACATCAATC 504

Db        516 ACGGCAGCCGAAGTGGCCGTATACCAGGACGTGTATGCTGTACATGCACCAACATCGCTG 575

Qy        505 TACCACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTGACACGACCCCT 564

Db        576 TACCATCAGGCGATGAAATGTGTCAGAACGGCGTATTGGATTGGGTTTGACACCACCCCG 635

Qy        565 TTTATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACTGGGCTGACGAG 624

Db        636 TTTATGTTTGACGCGCTAGCAGGCGCGTATCCAACCTACGGCACAAACTGGGCCGACGAG 695

Qy        625 AGAGTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGGAGAGCAGGCTT 684

Db        696 CAGGTGTTACAGGCCAGGAACATAGGACTGTGTGCAGCATCCTTGACTGAGGGAAGACTC 755

Qy        685 GGAAAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGATCATATTCTCG 744

Db        756 GGCAAACTGTCCATTCTCCGCAAGAAGCAATTGAAACCTTGCGACACAGTCATGTTCTCG 815

Qy        745 GTTGGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGCATCTTCCAAAC 804

Db        816 GTAGGATCTACATTGTACACTGAGAGCAGAAAGCTACTGAGGAGCTGGCACTTACCCTCC 875

Qy        805 GTGTTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTGTCAGCTGT 864

Db        876 GTATTCCACCTGAAAGGTAAACAATCCTTTACCTGTAGGTGCGATACCATCGTATCATGT 935

Qy        865 GAAGGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAGTTGAGAAC 924

Db        936 GAAGGGTACGTAGTTAAGAAAATCACTATGTGCCCCGGCCTGTACGGTAAAACGGTAGGG 995
```


Qy 925 TTGGCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATACGCTGCGC 984
Db 996 TACGCCGTGACGTATCACGCGGAGGGATTCCTAGTGTGCAAGACCACAGACACTGTCAAA 1055
Qy 985 GGCGAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCGATCAGATG 1044
Db 1056 GGAGAAAGAGTCTCATTCCCTGTATGGACCTACGTCCCCTCAACCATCTGTGATCAAATG 1115
Qy 1045 ACAGGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTATTGGTTGGGCTC 1104
Db 1116 ACTGGCATACTAGCGACCGACGTCACACCGGAGGACGCACAGAAGTTGTTAGTGGGATTG 1175
Qy 1105 AACCAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAATGCAGAACTAT 1164
Db 1176 AATCAGAGGATAGTTGTGAACGGAAGAACACAGCGAAACACTAACACGATGAAGAACTAT 1235
Qy 1165 CTATTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATCGTGCCGACTTG 1224
Db 1236 CTGCTTCCGATTGTGGCCGTCGCATTTAGCAAGTGGGCGAGGGAATACAAGGCAGACCTT 1295
Qy 1225 GACGACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCTGCTGCTGGGCT 1284
Db 1296 GATGATGAAAAACCTCTGGGTGTCCGAGAGAGGTCACTTACTTGCTGCTGCTTGTGTGCA 1355
Qy 1285 TTCAAGACCCAGAAAATCACATCCATCTACAAGAAGCCTGGTACGCAAACAATTAAGAAA 1344
Db 1356 TTTAAAACGAGGAAGATGCACACCATGTACAAGAAACCAGACACCCAGACAATAGTGAAG 1415
Qy 1345 GTACCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCACGGGCTCGATATG 1404
Db 1416 GTGGCTTCAGAGTTTAACTCGTTCGTCATCCCGAGCCTATGGTCTACAGGCCTCGCAATC 1475
Qy 1405 GGCTTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACTGTCAAACCCGCACCGGCTATT 1464
Db 1476 CCAGTCAGATCACGCATTAAGATGCTTTTGGCCAAGAAGACCAAGC---GAGAGTTAATA 1532
Qy 1465 ACAATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAGAAGTGGCTGCA 1524
Db 1533 CCTGTTCTCGACGCGTCGTCAGCCAGGGATGCTGAACAAGAGGAGAAGGAGATGTTGGAG 1592
Qy 1525 GCGGAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAAAAGAGACCGTA 1584
Db 1593 GCCGAGCTGACTAGAGAAGCCTTACCACCCCTCGTCCCCATCGCGCCGGCGGAGACGGGA 1652
Qy 1585 GAGGCAGAAGTAGACCTCATTATGCAAGAG-----GCAGGAGCAGGTAGCGTGGAGACA 1638
Db 1653 GTCGTCGACGTCGACGTTGAAGAACTAGAGTATCACGGAGGTGCAGGGGTCGTGGAAACA 1712
Qy 1639 CCACGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAGATTGGGTCTTACGCT 1698
Db 1713 CCTCGCAGCGCGTTGAAAGTCACCGCACAGCCGAACGACGTACTACTAGGAAATTACGTA 1772
Qy 1699 ATACTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGTATCCACCCATTGGCG 1758
Db 1773 GTTCTGTCCCCGCAGACCGTGCTCAAGAGCTCCAAGTTGGCCCCCGTGCACCCTCTAGCA 1832
Qy 1759 GAACAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATACAAAGTCGAGCCATAC 1818
Db 1833 GAGCAGGTGAAAATAATAACACATAACGGTAGGGCCGGCGGTTACCAGGTCGACGGATAT 1892
Qy 1819 CACGGTAAGGTCATTGTACCAGAAGGGACGGCGGTCCCTGTTCAAGACTTCCAGGCATTG 1878
Db 1893 GACGGCAGGGTCCTACTACCATGTGGATCGGCCATTCCGGTCCCTGAGTTTCAAGCTTTG 1952
Qy 1879 AGTGAGAGCGCTACGATCGTTTTCAACGAGAGGGAGTTCGTAAACAGATACCTGCACCAC 1938
Db 1953 AGCGAGAGCGCCACTATGGTGTACAACGAAAGGGAGTTCGTCAACAGGAAACTATACCAT 2012
Qy 1939 ATCGCAATCAACGGAGGAGCGCTAAACACTGACGAAGAGTACTATAAGACTGTAAAGACT 1998
Db 2013 ATTGCCGTTCACGGACCGTCGCTGAACACCGACGAGGAGAACTACGAGAAAGTCAGAGCT 2072
Qy 1999 CAGGACACAGACTCAGAATACGTCTTCGATATTGACGCACGAAAGTGTGTTAAGCGAGAA 2058
Db 2073 GAAAGAACTGACGCCGAGTACGTGTTCGACGTAGATAAAAAATGCTGCGTCAAGAGAGAG 2132
Qy 2059 GACGCAGGTCCCTTGTGCCTAACCGGTGATCTGGTAGATCCACCATTTCACGAGTTTGCG 2118
Db 2133 GAAGCGTCGGGTTTGGTGTTGGTGGGAGAGCTAACCAACCCCCCGTTCCATGAATTCGCC 2192
Qy 2119 TACGAGAGTCTCAAGACACGACCAGCAGCACCTCACAAAGTCCCAACCATCGGAGTCTAT 2178
Db 2193 TACGAAGGGCTGAAGATCAGGCCGTCGGCACCATATAAGACTACAGTAGTAGGAGTCTTT 2252
Qy 2179 GGAGTGCCAGGTTCAGGTAAATCTGGAATCATCAAAAGCGCTGTGACTAAGAAAGATCTG 2238
Db 2253 GGGGTTCCGGGATCAGGCAAGTCTGCTATTATTAAGAGCCTCGTGACCAAACACGATCTG 2312
Qy 2239 GTTGTGAGTGCGAAGAAGGAAAACTGCGCAGAAATCATCAGGGATGTAAGGAGGATGAGA 2298
Db 2313 GTCACCAGCGGCAAGAAGGAGAACTGCCAGGAAATAGTTAACGACGTGAAGAAGCACCGC 2372
Qy 2299 CGTATGGATGTTGCTGCTAGGACTGTCGATTCAGTGCTTCTAAATGGGGTTAAGCACCCC 2358
Db 2373 GGGAAGGGGACAAGTAGGGAAAACAGTGACTCCATCCTGCTAAACGGGTGTCGTCGTGCC 2432
Qy 2359 GTTAACACTCTGTACATTGATGAGGCATTTGCCTGCCATGCAGGGACGCTGCTGGCACTG 2418
Db 2433 GTGGACATGCTATATGTGGACGAGGCTTTCGCTTGCCATTCCGGTACTCTGCTGGGCCTA 2492
Qy 2419 ATTGCCATCGTCAAACCT---AAGAAAGTGGTATTGTGCGGGGACCCAAAACAATGCGGC 2475
Db 2493 ATTGCTCTTGTTAAACCTCGGAGCAAAGTGGTGTTATGCGGAGACCCCAAGCAATGCGGA 2552
Qy 2476 TTCTTTAACATGATGTGCCTGAAAGTACATTTTAACCATGACATATGCACTGAAGTGTAC 2535
Db 2553 TTCTTCAATATGATGCAGCTTAAGGTGAACTTCAACCACAACATCTGCACTGAAGTATGT 2612
Qy 2536 CATAAAAGCATCTCTAGGAGGTGCACACAGACTGTAACCGCCATCGTCTCCACGCTCTTC 2595
Db 2613 CATAAAAGTATATCCAGACGTTGCACGCGTCCAGTCACGGCCATCGTGTCTACGTTGCAC 2672
Qy 2596 TACGACAAGCGAATGAAGACGGTTAACCCATGTGCTGATAAAATCATCATAGATACCACA 2655
Db 2673 TACGGAGGCAAGATGCGCACGACCAACCCGTGCAACAAACCCATAATCATAGACACCACA 2732
Qy 2656 GGGACCACAAAGCCGCACAAAGATGATCTGATTCTAACCTGTTTCAGAGGATGGGTGAAA 2715
Db 2733 GGACAGACCAAGCCCAAGCCAGGAGACATCGTGTTAACATGCTTCCGAGGCTGGGCAAAG 2792
Qy 2716 CAGCTACAGATTGACTACAAAAATCACGAAATCATGACTGCGGCTGCATCGCAAGGACTT 2775
Db 2793 CAGCTGCAGTTGGACTACCGTGGACACGAAGTCATGACAGCAGCAGCATCTCAGGGCCTC 2852
Qy 2776 ACGCGGAAAGGCGTTTATGCTGTCAGGTACAAAGTCAACGAGAATCCACTCTACTCGCAG 2835
Db 2853 ACCCGCAAAGGGGTATACGCCGTAAGGCAGAAGGTGAATGAAAATCCCTTGTATGCCCCT 2912
Qy 2836 ACTTCTGAGCACGTGAACGTGTTACTTACACGCACAGAAAAACGCATTGTCTGGAAGACG 2895
Db 2913 GCGTCGGAGCACGTGAATGTACTGCTGACGCGCACTGAGGATAGGCTGGTGTGGAAAACG 2972
Qy 2896 CTAGCTGGTGATCCCTGGATAAAGACACTTACAGCTAAATATCCCGGGGATTTCACGGCT 2955
Db 2973 CTGGCCGGCGATCCCTGGATTAAGGTCCTATCAAACATTCCACAGGGTAACTTTACGGCC 3032
Qy 2956 TCATTGGACGACTGGCAGCGCGAACACGACGCCATTATGGCACGCGTTCTTGATAAGCCG 3015
Db 3033 ACATTGGAAGAATGGCAAGAAGAACACGACAAAATAATGAAGGTGATTGAAGGACCGGCT 3092
Qy 3016 CAGACAGCTGATGTGTTCCAGAATAAGGTGAACGTCTGCTGGGCGAAGGCTTTAGAGCCA 3075
Db 3093 GCGCCTGTGGACGCGTTCCAGAACAAAGCGAACGTGTGTTGTGCGAAAAGCCTGGTGCCT 3152
Qy 3076 GTCTTGGCCACGGCCAACATTGTGCTGACGAGACAGCAGTGGGAGAC---GTTGCACCCA 3132

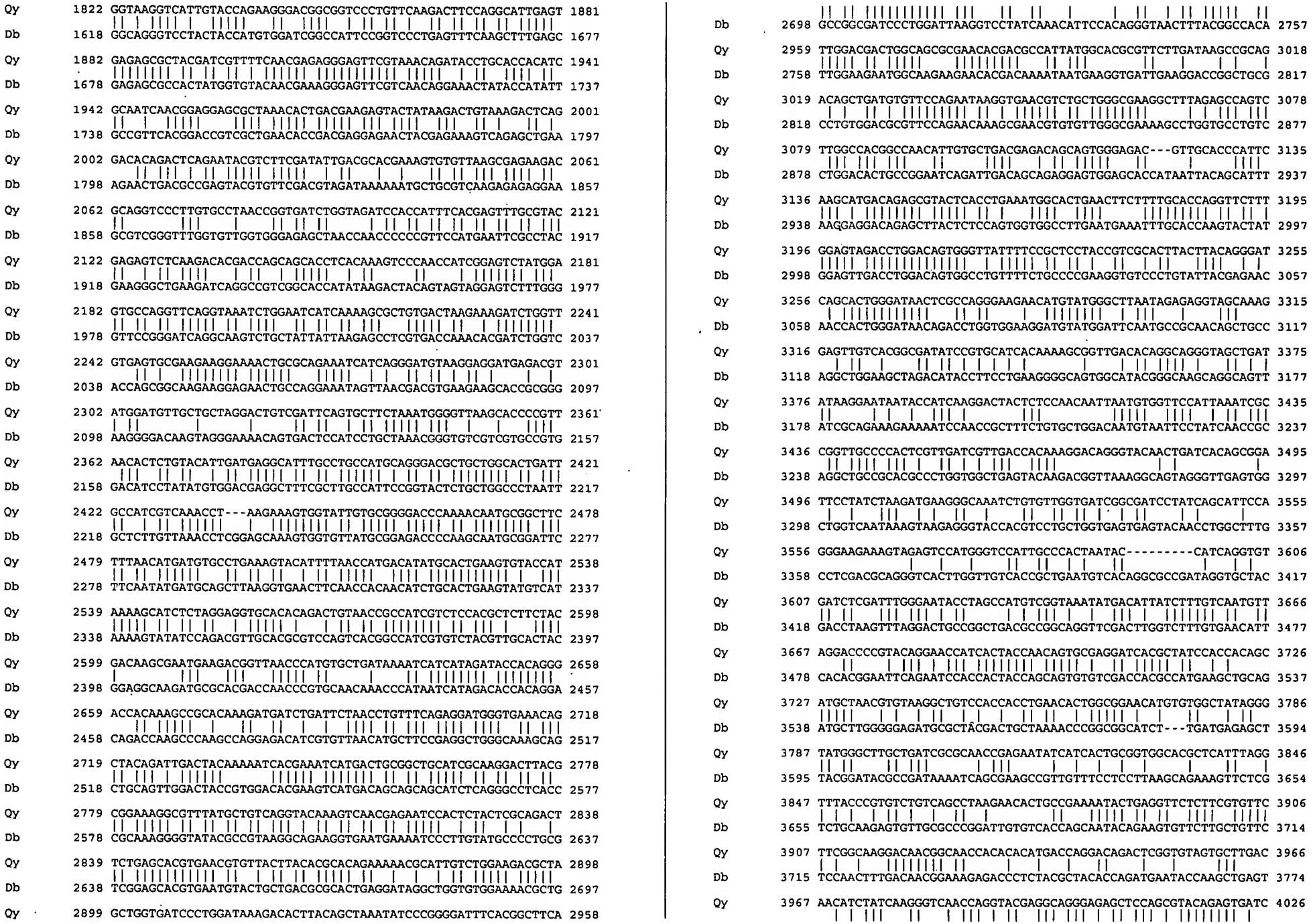
```
Db       3153 GTCCTGGACACTGCCGGAATCAGATTGACAGCAGAGGAGTGGAGCACCATAATTACAGCA 3212
Qy       3133 TTCAAGCATGACAGAGCGTACTCACCTGAAATGGCACTGAACTTCTTTTGCACCAGGTTC 3192
Db       3213 TTTAAGGAGGACAGAGCTTACTCTCCAGTGGTGGCCTTGAATGAAATTTGCACCAAGTAC 3272
Qy       3193 TTTGGAGTAGACCTGGACAGTGGGTTATTTTCCGCTCCTACCGTCGCACTTACTTACAGG 3252
Db       3273 TATGGAGTTGACCTGGACAGTGGCCTGTTTTCTGCCCCGAAGGTGTCCCTGTATTACGAG 3332
Qy       3253 GATCAGCACTGGGATAACTCGCCAGGGAAGAACATGTATGGGCTTAATAGAGAGGTAGCA 3312
Db       3333 AACAACCACTGGGATAACAGACCTGGTGGAAGGATGTATGGATTCAATGCCGGAACAGCT 3392
Qy       3313 AAGGAGTTGTCACGGCGATATCCGTGCATCACAAAAGCGGTTGACACAGGCAGGGTAGCT 3372
Db       3393 GCCAGGCTGGAAGCTAGACATACCTTCCTGAAGGGGCAGTGGCATACGGGCAAGCAGGCA 3452
Qy       3373 GATATAAGGAATAATACCATCAAGGACTACTCTCCAACAATTAATGTGGTTCCATTAAAT 3432
Db       3453 GTTATCGCAGAAAGAAAAATCCAACCGCTTTCTGTGCTGGACAATGTAAATCCTATCAAC 3512
Qy       3433 CGCCGGTTGCCCCACTCGTTGATCGTTGACCACAAAGGACAGGGTACAACTGATCACAGC 3492
Db       3513 CGCAGGCTGCCGCACGCCCTGGTGGCTGAGTACAAGACGGTTAAAGGCAGTAGGGTTGAG 3572
Qy       3493 GGATTCCTATCTAAGATGAAGGGCAAATCTGTGTTGGTGATCGGCGATCCTATCAGCATT 3552
Db       3573 TGGCTGGTCAATAAAGTAAGAGGCTACCACGTCCTGCTGGTGAGTGAGTACAACCTGGCT 3632
Qy       3553 CCAGGGAAGAAAGTAGAGTCCATGGGTCCATTGCCCACTAATAC---------CATCAGG 3603
Db       3633 TTGCCTCGACGCAGGGTCACTTGGTTGTCACCGCTGAATGTCACAGGCGCCGATAGGTGC 3692
Qy       3604 TGTGATCTCGATTTGGGAATACCTAGCCATGTCGGTAAATATGACATTATCTTTGTCAAT 3663
Db       3693 TACGACCTAAGTTTAGGACTGCCGGCTGACGCCGGCAGGTTCGACTTGGTCTTTGTGAAC 3752
Qy       3664 GTTAGGACCCCGTACAGGAACCATCACTACCAACAGTGCGAGGATCACGCTATCCACCAC 3723
Db       3753 ATTCACACGGAATTCAGAATCCACCACTACCAGCAGTGTGTCGACCACGCCATGAAGCTG 3812
Qy       3724 AGCATGCTAACGTGTAAGGCTGTCCACCACCTGAACACTGGCGGAACATGTGTGGCTATA 3783
Db       3813 CAGATGCTTGGGGGAGATGCGCTACGACTGCTAAAACCCGGCGGCATCT---TGATGAGA 3869
Qy       3784 GGGTATGGGCTTGCTGATCGCGCAACCGAGAATATCATCACTGCGGTGGCACGCTCATTT 3843
Db       3870 GCTTACGGATACGCCGATAAAATCAGCGAAGCCGTTGTTTCCTCCTTAAGCAGAAAGTTC 3929
Qy       3844 AGGTTTACCCGTGTCTGTCAGCCTAAGAACACTGCCGAAAATACTGAGGTTCTCTTCGTG 3903
Db       3930 TCGTCTGCAAGAGTGTTGCGCCCGGATTGTGTCACCAGCAATACAGAAGTGTTCTTGCTG 3989
Qy       3904 TTCTTCGGCAAGGACAACGGCAACCACACACATGACCAGGACAGACTCGGTGTAGTGCTT 3963
Db       3990 TTCTCCAACTTTGACAACGGAAAGAGACCCTCTACGCTACACCAGATGAATACCAAGCTG 4049
Qy       3964 GACAACATCTATCAAGGGTCAACCAGGTACGAGGCAGGGAGAGCTCCAGCGTACAGAGTG 4023
Db       4050 AGTGCCGTGTATGCCGGAGAAGGCATGCACACGGCCGGGTGTGCACCATCCTACAGAGTT 4109
Qy       4024 ATCAGAGGTGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGCAAA 4083
Db       4110 AAGAGAGCAGACATAGCCACGTGCACAGAAGCGGCTGTGGTTAACGGAGCTAACGCCCGT 4169
Qy       4084 GGTCAACCAGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTTGAT 4143
Db       4170 GGAACTGTAGGGGATGGCGTATGCAGGGCCGTGGCGAAGAAATGGCCGTCAGCCTTTAAG 4229
Qy       4144 AGACAGCCAATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACCG---CTCATCATA 4200
Db       4230 GGAGCAGCAACACCAGTGGGCACAATTAAAACAGTCATGTGCGGCTCGTACCCCGTCATC 4289
Qy       4201 CATGCTGTAGGACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTAAGCTCGCA 4260
Db       4290 CACGCTGTAGCGCCTAATTTCTCTGCCACGACTGAAGCGGAAGGGGACCGCGAATTGGCC 4349
Qy       4261 GCTGCCTACATGAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTA 4320
Db       4350 GCTGTCTACCGGGCAGTGGCCGCCGAAGTAAACAGACTGTCACTGAGCAGCGTAGCCATC 4409
Qy       4321 CCGCTACTGTCAACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCAT 4380
Db       4410 CCGCTGCTGTCCACAGGAGTGTTCAGCGGCGGAAGAGATAGGCTGCAGCAATCCCTCAAC 4469
Qy       4381 CACCTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAA 4440
Db       4470 CATCTATTCACAGCAATGGACGCCACGGACGCTGACGTGACCATCTACTGCAGAGACAAA 4529
Qy       4441 CAATGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATTCTGGAT 4500
Db       4530 AGTTGGGAGAAGAAAATCCAGGAAGCCATTGACATGAGGACGGCTGTGGAGTTGCTCAAT 4589
Qy       4501 GATGACAAGCCAGTAGACATTGACTTGGTCAGGGTCCACCCAAACAGCTCTTTGGCAGGC 4560
Db       4590 GATGACGTGGAGCTGACCACAGACTTGGTGAGAGTGCACCCGGACAGCAGCCTGGTGGGT 4649
Qy       4561 AGACCAGGTTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTGGAAGGTACACGATTC 4620
Db       4650 CGTAAGGGCTACAGTACCACTGACGGGTCGCTGTACTCGTACTTTGAAGGTACGAAATTC 4709
Qy       4621 CATCAGACCGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCCAACAAATCTGAGGCT 4680
Db       4710 AACCAGGCTGCTATTGATATGGCAGAGATACTGACGTTGTGGCCCAGACTGCAAGAGGCA 4769
Qy       4681 AATGAGCAGATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGC 4740
Db       4770 AACGAACAGATATGCCTATACGCGCTGGGCGAAACAATGGACAACATCAGATCCAAATGT 4829
Qy       4741 CCAGTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTAC 4800
Db       4830 CCGGTGAACGATTCCGATTCATCAACACCTCCCAGGACAGTGCCCTGGCTGTGCCGCTAC 4889
Qy       4801 GCTATGACGGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTA 4860
Db       4890 GCAATGACAGCAGAACGGATCGGCCGGCTTAGGTCACACCAAGTTAAAAGCATGGTGGTT 4949
Qy       4861 TGCTCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGC 4920
Db       4950 TGCTCATCTTTTCCCCTCCCGAAATACCATGTAGATGGGGTGCAGAAGGTAAAGTGCGAG 5009
Qy       4921 AAACCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGCGGAA 4980
Db       5010 AAGGTTCTCCTGTTCGACCCGACGGTACCTTCAGTGGTTAGTCGGCGGAAGTATGCCGCA 5069
```

RESULT 15
AAX78129
ID AAX78129 standard; DNA; 8010 BP.
XX
AC AAX78129;
XX
DT 20-AUG-1999 (first entry)
XX
DE Semliki Forest virus replicon fragment.
XX
KW Alphavirus; replicon; expression vector; heterologous splice site;
KW RNA splicing; viral RNA replication; antigen; vaccination; therapy; ss.
XX
OS Semliki forest virus.
XX
PN WO9925859-A1.
XX
PD 27-MAY-1999.

XX
PF 13-NOV-1998; 98WO-CA01065.
XX
PR 14-NOV-1997; 97US-0065793.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Parrington M;
XX
DR WPI; 1999-385140/32.
XX
PT New vector (EV) used to express an antigen for vaccination, e.g.
PT against human immune deficiency virus
XX
PS Disclosure; Fig 3A-F; 60pp; English.
XX
CC This invention describes the construction of a novel expression vector
CC (EV) containing an alphavirus replicon having a heterologous splice site.
CC The expression vector contains (i) a DNA molecule (I) complementary to
CC at least part of an alphavirus RNA genome, and (ii) inserted into a
CC region of (I) that is not essential for its replication, a heterologous
CC DNA (II), under control of a promoter. (I) includes at least one
CC heterologous splice site (HSS) to prevent aberrant RNA splicing of the
CC alphavirus and is the complement of the complete alphavirus genomic
CC region essential for replication of viral RNA. The expression vector is
CC used to express (II) in humans or animals, e.g. to express an antigen
CC for vaccination (against human immune deficiency virus) or to produce a
CC therapeutically active protein or peptide. Introducing an HSS: (1) makes
CC it more likely that any splicing will occur at this site, rather than at
CC a cryptic splice site; (2) restores function of the alphavirus when
CC eliminated, and (3) may improve transport of RNA from the nucleus.
XX
SQ Sequence 8010 BP; 2213 A; 2036 C; 2158 G; 1603 T; 0 other;

Query Match 29.5%; Score 1473.8; DB 20; Length 8010;
Best Local Similarity 57.9%; Pred. No. 0;
Matches 2779; Conservative 0; Mismatches 1982; Indels 36; Gaps 8;

```
Qy        214 ATTGGAAGTGCGCCCGTCAGACATGCACATTCCAATCACCGCTATCATTGTATCTGCCCT 273
              || || ||||||||   |||         || |  |||   || || ||  | ||||||
Db          1 ATCGGCAGTGCGCCTTCCAGGAGAATGATGTCTACGCACAAATACCACTGCGTATGCCCT 60

Qy        274 ATGATAAGCGCTGAAGACCCGGACAGACTACAACGGTATGCAGAAAGACTTAAGAAAA-- 331
              |||   ||||| |||||||| || ||  ||   |  | || ||| | | |||
Db         61 ATGCGCAGCGCAGAAGACCCCGAAAGGCTCGATAGCTACGCAAAGAAACTGGCAGCGGCC 120

Qy        332 ----GTGACATTACCGACAAGAACATAGCCTCTAAGGCGGCAGACCTGCTGGAAGTCATG 387
                  |  |  |    || |   | || ||     ||      | |||||||  |   ||||||
Db        121 TCCGGGAAGGTGCTGGATAGAGAGATCGCAGGAAAAATCACCGACCTGCAGACCGTCATG 180

Qy        388 TCAACACCAGACGCAGAGACTCCATCTCTGTGTATGCACACAGACGCCACGTGTAGGTAC 447
              | || |||||||| ||  ||||  |  | ||  |||| ||||||| ||||||| |
Db        181 GCTACGCCAGACGCTGAATCTCCTACCTTTTGCCTGCATACAGACGTCACGTGTCGTACG 240

Qy        448 TTTGGAAGTGTAGCAGTATACCAAGATGTGTACGCAGTCCATGCACCGACATCAATCTAC 507
                |    || || |||||||| || ||||| || || |||||||| |||||   | |||
Db        241 GCAGCCGAAGTGGCCGTATACCAGGACGTGTATGCTGTACATGCACCAACATCGCTGTAC 300

Qy        508 CACCAGGCGCTTAAAGGAGTTAGGACAATTTACTGGATAGGCTTTGACACGACCCCTTTT 567
              || |||||| | ||||| || || ||    || ||||| || |||||||| ||||| |||
Db        301 CATCAGGCGATGAAAGGTGTCAGAACGGCGTATTGGATTGGGTTTGACACCACCCCGTTT 360

Qy        568 ATGTACAAAAACATGGCAGGTTCCTACCCTACTTACAACACGAACTGGGCTGACGAGAGA 627
              ||||   |     | |||||  | || || || |||  ||| |||||||| ||||||
Db        361 ATGTTTGACGCGCTAGCAGGCGCGTATCCAACCTACGCCACAAACTGGGCCGACGAGCAG 420

Qy        628 GTATTGGAAGCACGTAACATTGGCCTCGGTAACTCAGATCTTCAGGAGAGCAGGCTTGGA 687
              || ||  | ||  | ||||| || ||  ||     ||     |    ||| | || || ||
Db        421 GTGTTACAGGCCAGGAACATAGGACTGTGTGCAGCATCCTTGACTGAGGGAAGACTCGGC 480

Qy        688 AAACTCTCAATCCTTAGGAAGAAGAGGCTCCAACCTACTAATAAGATCATATTCTCGGTT 747
              |||| || || ||  | ||||||    |  |||||    | |   |||| ||||||||
Db        481 AAACTGTCCATTCTCCGCAAGAAGCAATTGAAACCTTGCGACACAGTCATGTTCTCGGTA 540

Qy        748 GGTTCAACAATCTACACAGAAGATAGATCACTGTTACGTAGCTGGCATCTTCCAAACGTG 807
              || || ||| | ||||| ||    |||   ||  |  | ||||||||  | ||   |||
Db        541 GGATCTACATTGTACACTGAGAGCAGAAAGCTACTGAGGAGCTGGCACTTACCCTCCGTA 600

Qy        808 TTCCACTTGAAAGGAAAGTCTAACTTCACAGGTAGATGTGGGACCATTGTCAGCTGTGAA 867
              |||||| ||||||| ||      ||| ||  |||| || |  ||||| ||     ||||||
Db        601 TTCCACCTGAAAGGTAAACAATCCTTTACCTGTAGGTGCGATACCATCGTATCATGTGAA 660

Qy        868 GGGTACGTCATCAAAAAGATAACGATCAGCCCAGGACTATACGGTAAAGTTGAGAACTTG 927
              |||||||  | || || || || ||  |||| || || |||||||||   |     |
Db        661 GGGTACGTAGTTAAGAAAATCACTATGTGCCCCGGCCTGTACGGTAAAACGGTAGGGTAC 720

Qy        928 GCGTCCACAATGCATCGCGAGGGTTTCTTGAGTTGCAAAGTCACAGATACGCTGCGCGGC 987
              ||    ||    ||    ||||| ||| |    |||||   |||||| ||  |    ||
Db        721 GCCGTGACGTATCACGCGGAGGGATTCCTAGTGTGCAAGACCACAGACACTGTCAAAGGA 780

Qy        988 GAGAGGGTTTCTTTTGCTGTGTGTACGTATGTACCAGCCACACTTTGCGATCAGATGACA 1047
              || || || || ||  |||| || || || || ||  | ||  | || ||||| |||||
Db        781 GAAAGAGTCTCATTCCCTGTATGCACCTACGTCCCCTCAACCATCTGTGATCAAATGACT 840

Qy       1048 GGGATTCTGGCAACTGACGTTAGTGTGGATGACGCACAAAAACTATTGGTTGGGCTCAAC 1107
              || || || || || ||||| |     ||| |||||||| ||  | || || ||  | ||
Db        841 GGCATACTAGCGACCGACGTCACACCGGAGGACGCACAGAAGTTGTTAGTGGGATTGAAT 900

Qy       1108 CAAAGGATTGTCGTCAATGGTAGGACGCAAAGAAATACTAACACAATGCAGAACTATCTA 1167
              || ||||| || || || || || || || ||  |||| |||||||| ||| ||||||||||
Db        901 CAGAGGATAGTTGTGAACGGAAGAACACAGCGAAACACTAACACGATGAAGAACTATCTG 960

Qy       1168 TTACCAGTGGTCGCCCAGGCGTTTTCCAGGTGGGCGCGTGAACATCGTGCCGACTTGGAC 1227
              | ||  | || |||   || |||  || |||||| | |||  |    || ||| | ||
Db        961 CTTCCGATTGTGGCCGTCGCATTTAGCAAGTGGGCGAGGGAATACAAGGCAGACCTTGAT 1020

Qy       1228 GACGAGAAAGAACTAGGGGTGCGGGAGCGCACTCTTACTATGGGCTGCTGCTGGGCTTTC 1287
              || || |||   || || || || || ||| |  | ||||||    ||||||  ||||| ||
Db        1021 GATGAAAAACCTCTGGGTGTCCGAGAGAGGTCACTTACTTGCTGCTGCTTGTGGGCATTT 1080

Qy       1288 AAGACCCAGAAAATCACATCCATCTACAAGAAGCCTGGTACGCAAACAATTAAGAAAGTA 1347
              || ||   ||| ||      |||| |||||||| || |  || || |||||   ||| ||
Db        1081 AAAACGAGGAAGATGCACACCATGTACAAGAAACCAGACACCCAGACAATAGTGAAGGTG 1140

Qy       1348 CCTGCCGTCTTTGACTCATTTGTGATTCCACGCCTTACCAGCCACGGGCTCGATATGGGC 1407
              ||| | |  ||| |||| || || || ||  ||||          || ||||  ||
Db        1141 CCTTCAGAGTTTAACTCGTTCGTCATCCCGAGCCTATGGTCTACAGGCCTCGCAATCCCA 1200

Qy       1408 TTCCGCCGTAGGCTCAAGCTGCTGCTTGAACCAACTGTCAAACCCGCACCGGCTATTACA 1467
              || |     |  | ||| ||||  | |      |   ||| |  |   |  ||  |
Db        1201 GTCAGATCACGCATTAAGATGCTTTTGGCCAAGAAGACCAAGCGAGA---GTTAATACCT 1257

Qy       1468 ATGGCCGATGTGGAGCATCTGCGTGGCTTACAGCAAGAAGCTGAAGAAGTGGCTGCAGCG 1527
              |   ||| | |  |        | |      | |||||  |  | |||  |  |  ||
Db        1258 GTTCTCGACGCGTCGTCAGCCAGGGATGCTGAACAAGAGGAGAAGGAGAGGTTGGAGGCC 1317

Qy       1528 GAAGAGATCAGAGAAGCCCTGCCACCCTTGCTCCCTGAAATAGAAAAAGAGACCGTAGAG 1587
              ||  ||  |||||||||  | ||||||  |  ||||                |||||  | ||
Db        1318 GAGCTGACTAGAGAAGCCTTACCACCCCTCGTCCCCATCGCGCCGGCGGAGACGGGAGTC 1377

Qy       1588 GCAGAAGTAGACCTCATTATGCAAGAG------GCAGGAGCAGGTAGCGTGGAGACACCA 1641
              |  || || |||  |       | ||||      |||||  |||||   ||||||  |||||
Db        1378 GTCGACGTCGACGTTGAAGAACTAGAGTATCACGCAGGTGCAGGGGTCGTGGAAACACCT 1437

Qy       1642 CGAGGACACATCAGGGTGACAAGTTACCCAGGCGAAGAGAAGATTGGGTCTTACGCTATA 1701
              ||  |     | |  || ||      | ||   ||| |      | ||   ||||   |
Db        1438 CGCAGCGCGTTGAAAGTCACCGCACAGCCGAACGACGTACTACTAGGAAATTACGTAGTT 1497

Qy       1702 CTTTCACCCCAGGCGGTATTGAATAGTGAAAAACTGGCGTGTATCCACCCATTGGCGGAA 1761
              || || || ||| | ||  | || ||     ||  ||||      | |||||  | || ||
Db        1498 CTGTCCCCGCAGACCGTGCTCAAGAGCTCCAAGTTGGCCCCCGTGCACCCTCTAGCAGAG 1557

Qy       1762 CAAGTACTGGTAATGACTCACAAAGGTAGGGCAGGGAGATACAAAGTCGAGCCATACCAC 1821
              || ||      |||| || || || || ||||| ||  | ||| | |||||   ||| ||
Db        1558 CAGGTGAAAATAATAACACATAACGGGAGGGCCGGCGGTTACCAGGTCGACGGATATGAC 1617
```



```
Db       3775 GCCGTGTATGCCGGAGAAGCCATGCACACGGCCGGGTGTGCACCATCCTACAGAGTTAAG 3834

Qy       4027 AGAGGTGACATTAGCAAGAGCGCTGACCAAGCTATCGTTAATGCTGCTAATAGCAAAGGT 4086

Db       3835 AGAGCAGACATAGCCACGTGCACAGAAGCGGCTGTGGTTAACGCAGCTAACGCCCGTGGA 3894

Qy       4087 CAACCAGGTTCCGGAGTGTGCGGTGCACTGTACCGAAAATGGCCGGCTGCTTTTGATAGA 4146

Db       3895 ACTGTAGGGGATGGCGTATGCAGGGCCGTGGCGAAGAAATGGCCGTCAGCCTTTAAGGGA 3954

Qy       4147 CAGCCAATAGCTGTCGGGACGGCTAGACTTGTGAAGCACGAACCG---CTCATCATACAT 4203

Db       3955 GCAGCAACACCAGTGGGCACAATTAAAACAGTCATGTGCGGCTCGTACCCCGTCATCCAC 4014

Qy       4204 GCTGTAGGACCCAATTTTTCTAAGATGCCGGAACCGGAGGGCGACCTTAAGCTCGCAGCT 4263

Db       4015 GCTGTAGCGCCTAATTTCTCTGCCACGACTGAAGCGGAAGGGGACCGCGAATTGGCCGCT 4074

Qy       4264 GCCTACATGAGCATAGCGTCCATCGTCAACGCTGAGCGGATTACAAAAATATCAGTACCG 4323

Db       4075 GTCTACCGGGCAGTGGCCGCCGAAGTAAACAGACTGTCACTGAGCAGCGTAGCCATCCCG 4134

Qy       4324 CTACTGTCAACCGGCATCTATTCTGGTGGCAAAGATCGAGTGATGCAATCATTGCATCAC 4383

Db       4135 CTGCTGTCCACAGGAGTGTTCAGCGGCGGAAGAGATAGGCTGCAGCAATCCCTCAACCAT 4194

Qy       4384 CTGTTCACTGCTTTCGACACTACGGATGCCGATGTCACCATATATTGCTTGGATAAACAA 4443

Db       4195 CTATTCACAGCAATGGACGCCACGGACGCTGACGTGACCATCTACTGCAGAGACAAAAGT 4254

Qy       4444 TGGGAGACCAGGATAATCGAGGCCATTCACCGCAAAGAAAGCGTCGAAATTCTGGATGAT 4503

Db       4255 TGGGAGAAGAAAATCCAGGAAGCCATTGACATGAGGACGGCTGTGGAGTTGCTCAATGAT 4314

Qy       4504 GACAAGCCAGTAGACATTGACTTGGTCAGGGTCCACCCAAACAGCTCTTTGGCAGGCAGA 4563

Db       4315 GACGTGGAGCTGACCACAGACTTGGTGAGAGTGCACCCGGACAGCAGCCTGGTGGGTCGT 4374

Qy       4564 CCAGGTTACTCCGTCAATGAGGGCAAGTTGTATTCATACCTGGAAGGTACACGATTCCAT 4623

Db       4375 AAGGGCTACAGTACCACTGACGGGTCGCTGTACTCGTACTTTGAAGGTACGAAATTCAAC 4434

Qy       4624 CAGACCGCCAAGGACATTGCCGAAATCCATGCAATGTGGCCCAACAAATCTGAGGCTAAT 4683

Db       4435 CAGGCTGCTATTGATATGGCAGAGATACTGACGTTGTGGCCCAGACTGCAAGAGGCAAAC 4494

Qy       4684 GAGCAGATTTGCTTGTACATCCTGGGGGAGAGTATGTCCAGCATCCGCTCCAAATGCCCA 4743

Db       4495 GAACAGATATGCCTATACGCGCTGGGCGAAACAATGGACAACATCAGATCCAAATGTCCG 4554

Qy       4744 GTAGAGGAGTCAGAGGCGTCTGCTCCACCTCACACACTTCCATGCCTGTGTAATTACGCT 4803

Db       4555 GTGAACGATTCCGATTCATCAACACCTCCCAGGACAGTGCCCTGCCTGTGCCGCTACGCA 4614

Qy       4804 ATGACGGCTGAGCGCGTATACAGGTTGCGCTCTGCGAAGAAAGAACAGTTCGCCGTATGC 4863

Db       4615 ATGACAGCAGAACGGATCGCCCGCCTTAGGTCACACCAAGTTAAAAGCATGGTGGTTTGC 4674

Qy       4864 TCATCATTCCTGTTGCCGAAGTACAGGATCACAGGCGTGCAGAAGCTACAGTGCAGCAAA 4923

Db       4675 TCATCTTTTCCCCTCCCGAAATACCATGTAGATGGGGTGCAGAAGGTAAAGTGCGAGAAG 4734

Qy       4924 CCAGTCCTGTTTTCAGGCGTCGTACCACCGGCTGTACACCCCAGGAAGTACGCGGAA 4980

Db       4735 GTTCTCCTGTTCGACCCGACGGTACCTTCAGTGGTTAGTCCGCGGAAGTATGCCGCA 4791


Search completed: November 15, 2003, 12:34:31
Job time : 1215 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 20:27:52 ; Search time 1505 Seconds
(without alignments)
11631.798 Million cell updates/sec

Title: US-10-023-649-1_COPY_5000_11484
Perfect score: 6485
Sequence: 1 caccgccagcaacgacaacc.....tattttgtttttaaaatttc 6485

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5877.6	90.6	11492	20	AAV74107	Western equine enc
2	2397.6	37.0	11663	19	AAV33475	South African Arbo
3	2396	36.9	11663	19	AAV33473	South African Arbo
4	2388.4	36.8	11703	19	AAV33476	Sindbis virus clon
5	2386.4	36.8	11717	19	AAV33474	Girdwood S.A.virus
6	2383.6	36.8	11703	21	AAC64506	Alphavirus SinDCCh
7	2382	36.7	11703	21	AAC64507	Alphavirus SinChir
8	2381.6	36.7	11663	18	AAT47668	South African Arbo
9	2380.4	36.7	11740	25	ABX81526	Sindbis virus cDNA
10	2370.8	36.6	16656	17	AAT30787	Alphavirus-based e
11	2370.6	36.6	11717	21	AAA70577	Sindbis-like virus
12	2369.2	36.5	16656	19	AAV60125	Representative euk
13	2369.2	36.5	16656	19	AAV42364	Representative euk
14	2369.2	36.5	16656	20	AAV70684	Representative euk
15	2369.2	36.5	16656	21	AAZ92765	Representative euk
16	2369.2	36.5	16656	21	AAZ92892	Representative euk
17	2369.2	36.5	16656	24	AAL38772	DNA of eukaryotic
18	2367.6	36.5	16656	24	ABK46238	Eukaryotic Layered
19	2364.4	36.5	16656	16	AAQ86154	Eukaryotic layered
20	2331	35.9	11626	21	AAA70607	Sindbis-like virus
21	2322.4	35.8	16958	25	AAL55269	DNA of expression
22	2225.2	34.3	11687	24	AAD25136	SinChironLP virus
23	2198.4	33.9	11684	24	AAD25135	SinDCchiron virus
24	2122.6	32.7	13905	20	AAX77358	Polynucleotide seq
25	2122.6	32.7	13905	21	AAA90390	Plasmid pTE5'2J DN
26	1738.8	26.8	11464	20	AAV74108	Venezuelan equine
27	1728.6	26.7	11459	21	AAA49442	Venezuelan equine
28	1650.8	25.5	7970	20	AAX77357	Polynucleotide seq
29	1650.8	25.5	7970	21	AAA90389	Plasmid pDH-EB DNA
30	1646.6	25.4	9542	20	AAX77359	Polynucleotide seq
31	1646.6	25.4	9542	21	AAA90391	987BBneo DNA seque
32	1415.4	21.8	11517	13	AAQ26021	pSP6-SFV4 RNA tran
33	926.2	14.3	12523	24	AAD29131	Venezuelan equine
34	924.6	14.3	12379	24	AAD29139	Human immunodefici
35	924.6	14.3	13584	24	AAD29141	Venezuelan equine
36	916.2	14.1	7479	24	AAD29132	Venezuelan equine
37	863.2	13.3	4003	20	AAV74109	Venezuelan equine
38	827.6	12.8	8000	25	ABX81525	Sindbis virus cDNA
39	824.4	12.7	8000	18	AAX59321	Sindbis virus vari
40	824.4	12.7	8000	20	AAX58571	Sindbis virus vari
41	824.4	12.7	8000	25	ABX81524	Sindbis virus cDNA
42	819.2	12.6	9951	20	AAX77354	Polynucleotide seq
43	819.2	12.6	9951	21	AAA90386	Sindbis expression
44	819.2	12.6	9951	22	AAD04742	Alphaviral vector
45	819.2	12.6	9951	24	ABV73174	Nucleotide sequenc

ALIGNMENTS

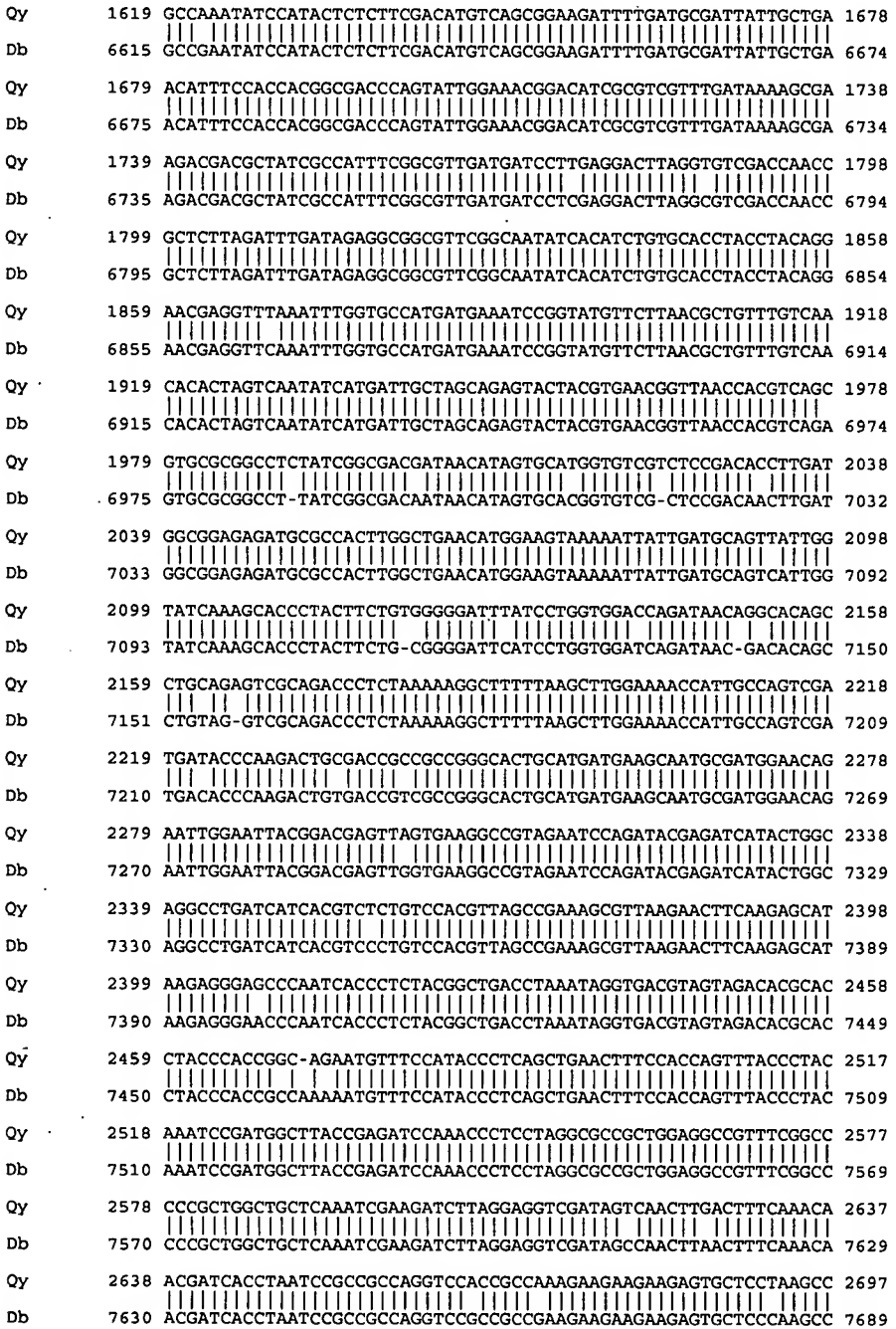
RESULT 1
AAV74107
ID AAV74107 standard; cDNA; 11492 BP.
XX
AC AAV74107;
XX
DT 12-APR-1999 (first entry)
XX
DE Western equine encephalitis virus cDNA.
XX
KW WEE virus; vaccine; ds.
XX
OS Western equine encephalitis virus.
XX
PN WO9853077-A1.
XX
PD 26-NOV-1998.
XX
PF 20-MAY-1998; 98WO-US10645.
XX
PR 16-DEC-1997; 97US-0991840.
PR 20-MAY-1997; 97US-0047162.
PR 24-JUL-1997; 97US-0053652.
XX
PA (REED-) REED ARMY INST RES WALTER.
XX
PI Crise BJ, Oberste MS, Parker MD, Schmura SM, Smith JF;
XX
DR WPI; 1999-045316/04.
XX
PT New DNA encoding infectious Western or Venezuelan equine

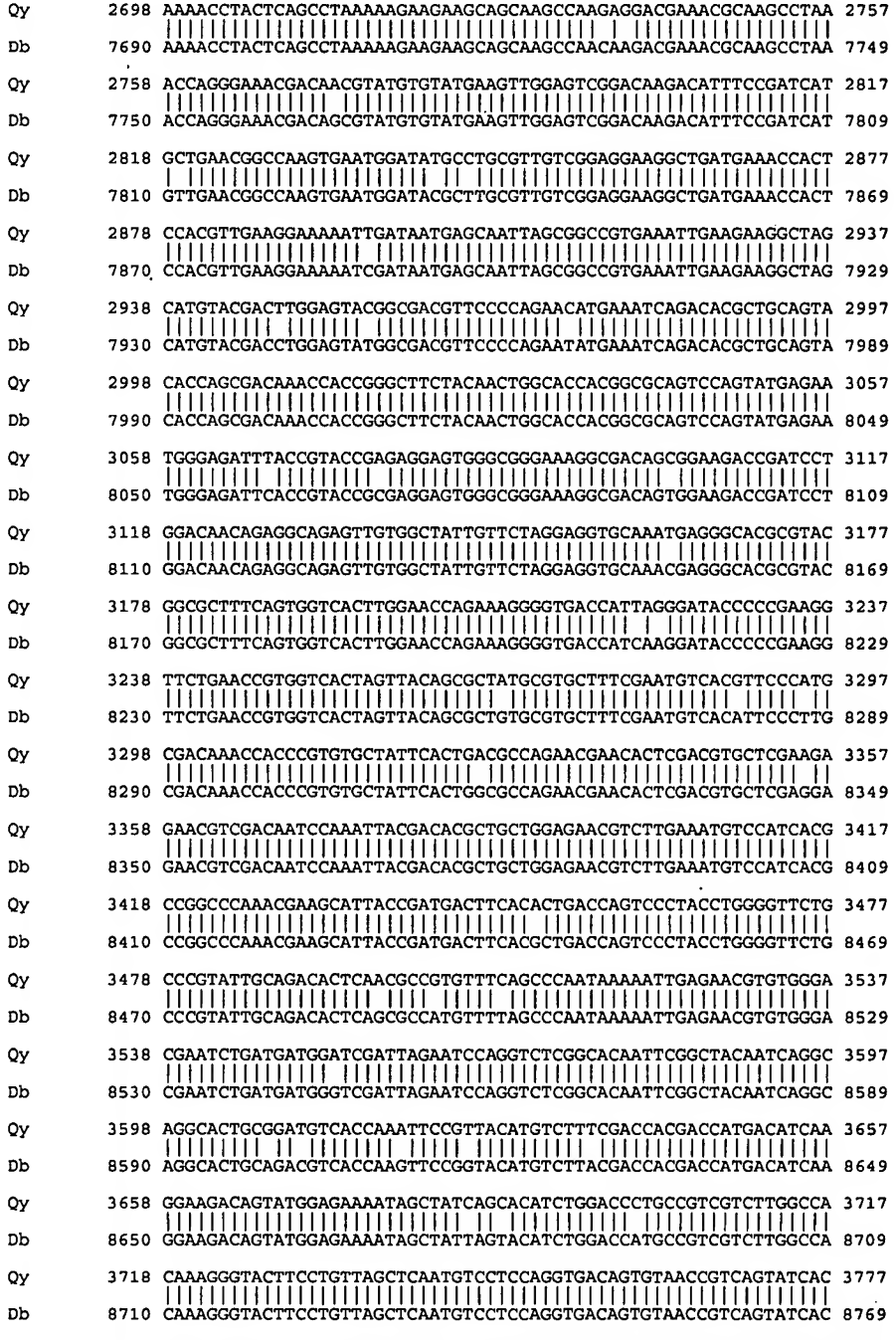
PT encephalitis virus genome - useful for the production of live or
PT attenuated vaccines for human or veterinary medicine
XX
PS Example 1; Page 54-67; 112pp; English.
XX
CC This cDNA sequence codes for an infectious western equine
CC encephalitis (WEE) virus RNA transcript. DNA representing the
CC entire genome was prepared by PCR using primers (see AAV74110-21)
CC based on partial genome sequences. 5' Sequences were obtained by
CC RACE. The full-length infectious clone is useful in the production
CC of virulent WEE virus, and for introducing and testing attenuating
CC mutations. Also new are: infectious or attenuated WEE RNA
CC transcripts and WEE viral particles; cDNA (see AAV74108) encoding an
CC infectious Venezuelan equine encephalitis (VEE) variant IE viral
CC genome; infectious or attenuated VEE RNA transcripts and viral
CC particles; an attenuated chimeric virus containing non-structural
CC sequences from a first alpha-virus (aV) and structural sequences
CC from a second aV, resulting in attenuation of the second aV; a method
CC for expressing a protein (especially an antigen to protect against
CC a pathogen) by cloning its gene into an attenuated WEE or VEE IE
CC replicon so that transcription of the replicon produces RNA able to
CC infect the cells in which protein is to be produced; methods for
CC diagnosing WEE and VEE infections; and polypeptides encoded by VEE
CC variant IIIA. Attenuated WEE and VEE are used in live or inactivated
CC vaccines, for use in human or veterinary medicine. Chimeric viruses
CC are also useful as vaccines, directed against the second aV which is
CC particularly Eastern equine encephalitis (EEE) virus or some variant
CC of VEE. WEE and VEE nucleic acids are used as primers and probes to
CC diagnose virus infections and to define natural variants, also for
CC production of protein antigens which can be used as diagnostic
CC reagents, to generate antibodies, and in vaccines. The attenuated
CC viruses are highly immunogenic and provide long-lasting protection.
XX
SQ Sequence 11492 BP; 3257 A; 2854 C; 2794 G; 2570 T; 17 other;

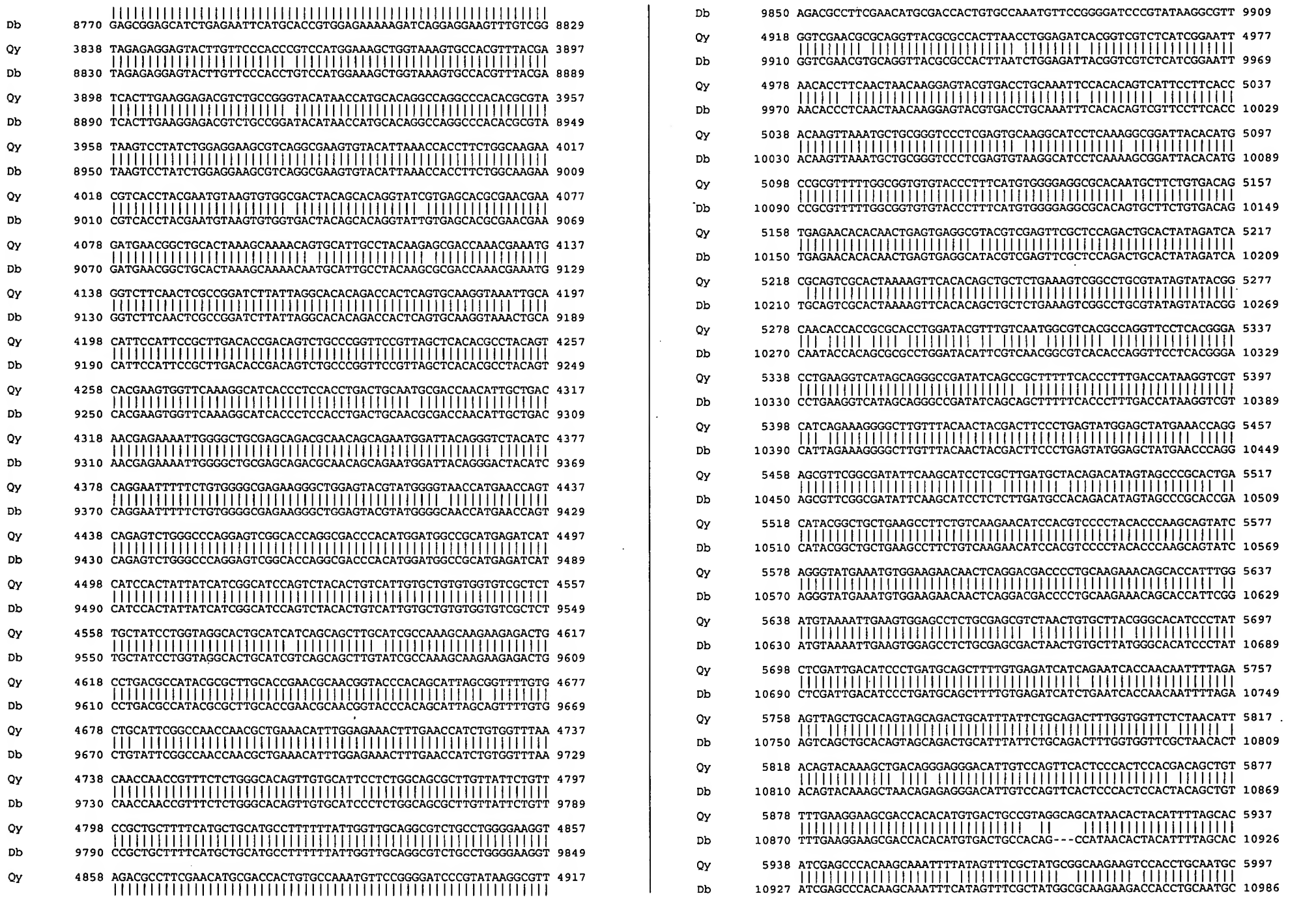
Query Match 90.6%; Score 5877.6; DB 20; Length 11492;
Best Local Similarity 96.1%; Pred. No. 0;
Matches 6238; Conservative 0; Mismatches 221; Indels 29; Gaps 20;

```
Qy          1 CACCGCCAGCAACGACAACCGTAATATGTGAACCAACTGTGCCAGAACGTATACCCAGTC 60
              ||  ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5012 CATCGCCAACAACGACAACCGTAATATGTGAACCAACTGTGCCAGAACGTATACCCAGTC 5071

Qy         61 CGGTGATTTCTAGAGCACCAAGTGCGGAATCACTGCTATCGCTTGGCGGCGTCTCGTTCT 120
              |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db       5072 CGGTGATTTCTAGAGCAGCAAGTGCGGAATCACTGCTATCGTTTGGCGGCGTCTCGTTCT 5131

Qy        121 CTAGCTCTGCCACACGCTCGTCAACCGCCTGGAGCGACTATGACAGGCGGTTTGTGGTTA 180
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       5132 CTAGCTCTGCCACACGCTCGTCAACTGCCTGGAGCGACTATGACAGGCGGTTTGTGGTTA 5191

Qy        181 CAGCTGATGTGCATCAAGCGAACACGTCTACGTGGAGCATCCCTAGTGCTCCTGGCTTGG 240
              |||||||  |||||||||||  |||||||||||||||||||||||||||||  |||||||||
Db       5192 CAGCTGACGTGCATCAAGCGGACACGTCTACGTGGAGCATCCCTAGCGCTCCTGACTTGG 5251

Qy        241 ACGTCCAGCTGCCTTCTGACGTCACTGATTCCCACTGGAGTATTCCAAGTGCATCAGGCT 300
              ||||||||||||||||||  |||||||||||||||||||||||||||  ||||||||||||
Db       5252 ACGTCCAGCTGCCTTCTGACGATACTGATTCCCACTGGAGTATTCCGAGTGCATCAGGCT 5311

Qy        301 TTGAAGTGAGAACACCATCTGTACAGGACCTAACTGCGGAGTGTGCGAAGCCTCGTGGAC 360
              |||||||||||||||| |||||||||||||||||||| |||||||||| ||||||||||||
Db       5312 TTGAAGTGAGAACACCGTCTGTACAGGACCTAACTGCAGAGTGTGCGAGGCCTCGTGGAC 5371

Qy        361 TGGCCGAAATAATGCAAGACTTCAATACTGCTCCTTTCCAGTTTCTTTCGGACTACAGAC 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5372 TGGCCGAAATAATGCAAGACTTCAATACGGCTCCTTTCCAGTTTCTTTCGGACCACAGAC 5431

Qy        421 CAGTACCGGCACCACGGAGACGCCCCATCCCATCACCTAGATCGACGGCTTCCGCACCTC 480
              |||||||  ||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       5432 CAGTACCAGCACCACGGAGACGCCCCATCCCATCACCTAGATCGACGGTTTCCGCACCTC 5491

Qy        481 CAGTTCCAAAGCCACGCAGGACTAAGTACCAACAACCACCAGGAGTCGCTAGAGCGATCT 540
              |||||||||||||||||| |||||||||||||||||||||||||||||||||| |||||||
```

```
Db       5492 CAGTTCCAAAGCCACGCAGAACTAAGTACCAACAACCACCAGGAGTCGCTAGGGCGATCT 5551

Qy .      541 CAGAAGCGGAGTTGGACGAGTACATCCGTCAACACTCCAACTGACGGTATGAAGCGGGAG 600
              |||||||||||  |||||||||||||||||||||||  |||||  |||||||||||||| ||||
Db       5552 CAGAAGCGGAGCTGGACGAGTACATCCGTCAACATTCCAATTGACGGTATGAAGC-GGAG 5610

Qy        601 CGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTCAACAGAAATCAGTACGTCAAT 660
              ||||||||   ||       ||   | | ||||||||||||||||||||||||||||||||||
Db       5611 CGTATATT---TCTCATCGAAACAGGCCAGGTCACCTTCAACAGAAATCAGTACGTCAAT 5667

Qy        661 GTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATGAGAAGTATTACGCCCCGCGCC 720
              ||||||||||||  ||||||||||  |||||||||||||||||||||||| |||||||||
Db       5668 GTAAACTACAAGACCCTATATTGGAACGGGCCGTCCATGAGAAGTATTA-GCCCCGCGCC 5726

Qy        721 TCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGCAATTATGCGCCTC-TGAAGGA 779
              |||||||||||||||||| |||||||||||||||||||||||||||||||| |||||||
Db       5727 TCGATCTCGAAAGAGAGAAGATGTTACAGAAGAAACTGCAATTATGCGCCTCNTGAAGGA 5786

Qy        780 AATAGAAGCAGGTATCAATCACGAAAAGTA-GAAAATATGAAAGCAATTACAGCGGAGCG 838
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db       5787 AATAGAAGCAGGTATCAATCACGAAAAGTANGAAAATATGAAAGCAATTACAGCGGAGCG 5846

Qy        839 ACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGAATCCTGTCGAGTGTTACAG 898
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5847 ACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGAATCCTGTCGAGTGTTACAG 5906

Qy        899 AGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTAACAGGTTTACATCTGCAGA 958
              |||||||||||||||||||||||||||||||||||||| |||| |||||| |||| |||||
Db       5907 AGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAGTTAATAGGTTTTCATCAGCAGA 5966

Qy        959 GGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATTACCCTACAGTAGCCAGTTA 1018
              || || || |||   |||||||||||| |||||||||||||||||||||||||||| |||
Db       5967 AGTGGCCGTCAAAGTTTGCAACTTAGTCATCCAAGAGAATTACCCTACAGTAGCCACTTA 6026

Qy       1019 TTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGGACGGCGCATCGTGCTGTCT 1078
              ||| |||||||||||||||||||||||||||||||||||||| ||||||| ||||||||| |
Db       6027 TTGCATAACAGATGAATACGATGCGTATCTTGACATGGTGTACGGCGC-TCGTGCTGT-T 6084

Qy       1079 AGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACCCAAAGAAGCATAGCTATTT 1138
              |||||||||  ||||  |   |      |||||||||||||||||||||||||||||| ||
Db       6085 AGATACAGCGCCTTTGTCCGC----TNACTGAGAAGCTACCCAAAGAAGCATAGCTACTT 6140

Qy       1139 GCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGAATACATTACAAAATGTATT 1198
              |||||||||||||||||||||||||||||| |||||||| ||||||||||||||||||||
Db       6141 GCAGCCAGAGATAAGATCAGCTGTCCCATCACCTATACAGAATACATTACAAAATGTATT 6200

Qy       1199 GGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGCGAGAATTACCTGTCTTAGA 1258
              ||||||||  |||||||||||  |||||||||||||||||||||||||||  |||||||  |
Db       6201 -GCTGCAGCCACTAAAAGGAACTGCAACGTTACCCAAATGCGAGAACTACCTGTTTT--A 6257

Qy       1259 TTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCATGCAATGATGAGTACTGGGA 1318
              ||||||||||| || ||||| |||||||||||||||||||||||||||||||||| |||||
Db       6258 TTCGGCGGCATTCAACGTTGAATGTTTCAAGAAATACGCATGCAATGATGAGTATTGGGA 6317

Qy       1319 TACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACGTTACGCAATATGTGACAAA 1378
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db       6318 TACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACGTTACGCAATACGTGACAAA 6377

Qy       1379 GCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTCATAATCTAAAACCGTTGCA 1438
              |||||||||||||||||||||||||||| || || || ||||| || ||||||||||||||
Db       6378 GCTGAAAGGGCCGAAAGCAGCAGCATTATTCGCAAAAACTCACAACCTAAAACCGTTGCA 6437

Qy       1439 GGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAGATGTCAAAGTTACTCCCGG 1498
              |||||||||||||||  |||| |||||||  ||| || ||||||||||||||||||||||
Db       6438 GGAGATACCAATGGACGAATTTGTCATGG-TCTNAAAAGAGATGTCAAAGTTACTCCCGG 6496

Qy       1499 CACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTCAGGCTGCAGATCCCCTTGC 1558
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db       6497 CACAAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTCAGGCTGCAGATCCTCTCTG 6556

Qy       1559 TACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTAGACTGAATGCGGTGCTTCT 1618
              |||||||||||||||||||||||||||| |||| |||||||||||||||| ||||||||
Db       6557 TACCGCTTACCTTTGCGGGATCCATCGAGAAT--GTCCGTAGACTGAATGCTGTGCTTCT 6614
```


Qy 1619 GCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATTTTGATGCGATTATTGCTGA 1678
Db 6615 GCCGAATATCCATACTCTCTTCGACATGTCAGCGGAAGATTTTGATGCGATTATTGCTGA 6674
Qy 1679 ACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCGCGTCGTTTGATAAAAGCGA 1738
Db 6675 ACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCGCGTCGTTTGATAAAAGCGA 6734
Qy 1739 AGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGGACTTAGGTGTCGACCAACC 1798
Db 6735 AGACGACGCTATCGCCATTTCGGCGTTGATGATCCTCGAGGACTTAGGCGTCGACCAACC 6794
Qy 1799 GCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACATCTGTGCACCTACCTACAGG 1858
Db 6795 GCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACATCTGTGCACCTACCTACAGG 6854
Qy 1859 AACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGTTCTTAACGCTGTTTGTCAA 1918
Db 6855 AACGAGGTTCAAATTTGGTGCCATGATGAAATCCGGTATGTTCTTAACGCTGTTTGTCAA 6914
Qy 1919 CACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTGAACGGTTAACCACGTCAGC 1978
Db 6915 CACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTGAACGGTTAACCACGTCAGA 6974
Qy 1979 GTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTGTCGTCTCCGACACCTTGAT 2038
Db 6975 GTGCGCGGCCT-TATCGGCGACAATAACATAGTGCACGGTGTCG-CTCCGACAACTTGAT 7032
Qy 2039 GGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAATTATTGATGCAGTTATTGG 2098
Db 7033 GGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAATTATTGATGCAGTCATTGG 7092
Qy 2099 TATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGGACCAGATAACAGGCACAGC 2158
Db 7093 TATCAAAGCACCCTACTTCTG-CGGGGATTCATCCTGGTGGATCAGATAAC-GACACAGC 7150
Qy 2159 CTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTGGAAAACCATTGCCAGTCGA 2218
Db 7151 CTGTAG-GTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTGGAAAACCATTGCCAGTCGA 7209
Qy 2219 TGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATGAAGCAATGCGATGGAACAG 2278
Db 7210 TGACACCCAAGACTGTGACCGTCGCCGGGCACTGCATGATGAAGCAATGCGATGGAACAG 7269
Qy 2279 AATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCAGATACGAGATCATACTGGC 2338
Db 7270 AATTGGAATTACGGACGAGTTGGTGAAGGCCGTAGAATCCAGATACGAGATCATACTGGC 7329
Qy 2339 AGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCGTTAAGAACTTCAAGAGCAT 2398
Db 7330 AGGCCTGATCATCACGTCCCTGTCCACGTTAGCCGAAAGCGTTAAGAACTTCAAGAGCAT 7389
Qy 2399 AAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGGTGACGTAGTAGACACGCAC 2458
Db 7390 AAGAGGGAACCCAATCACCCTCTACGGCTGACCTAAATAGGTGACGTAGTAGACACGCAC 7449
Qy 2459 CTACCCACCGGC-AGAATGTTTCCATACCCTCAGCTGAACTTTCCACCAGTTTACCCTAC 2517
Db 7450 CTACCCACCGCCAAAAATGTTTCCATACCCTCAGCTGAACTTTCCACCAGTTTACCCTAC 7509
Qy 2518 AAATCCGATGGCTTACCGAGATCCAAACCCTCCTAGGCGCCGCTGGAGGCCGTTTCGGCC 2577
Db 7510 AAATCCGATGGCTTACCGAGATCCAAACCCTCCTAGGCGCCGCTGGAGGCCGTTTCGGCC 7569
Qy 2578 CCCGCTGGCTGCTCAAATCGAAGATCTTAGGAGGTCGATAGTCAACTTGACTTTCAAACA 2637
Db 7570 CCCGCTGGCTGCTCAAATCGAAGATCTTAGGAGGTCGATAGCCAACTTAACTTTCAAACA 7629
Qy 2638 ACGATCACCTAATCCGCCGCCAGGTCCACCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCC 2697
Db 7630 ACGATCACCTAATCCGCCGCCAGGTCCGCCGCCGAAGAAGAAGAAGAGTGCTCCCAAGCC 7689

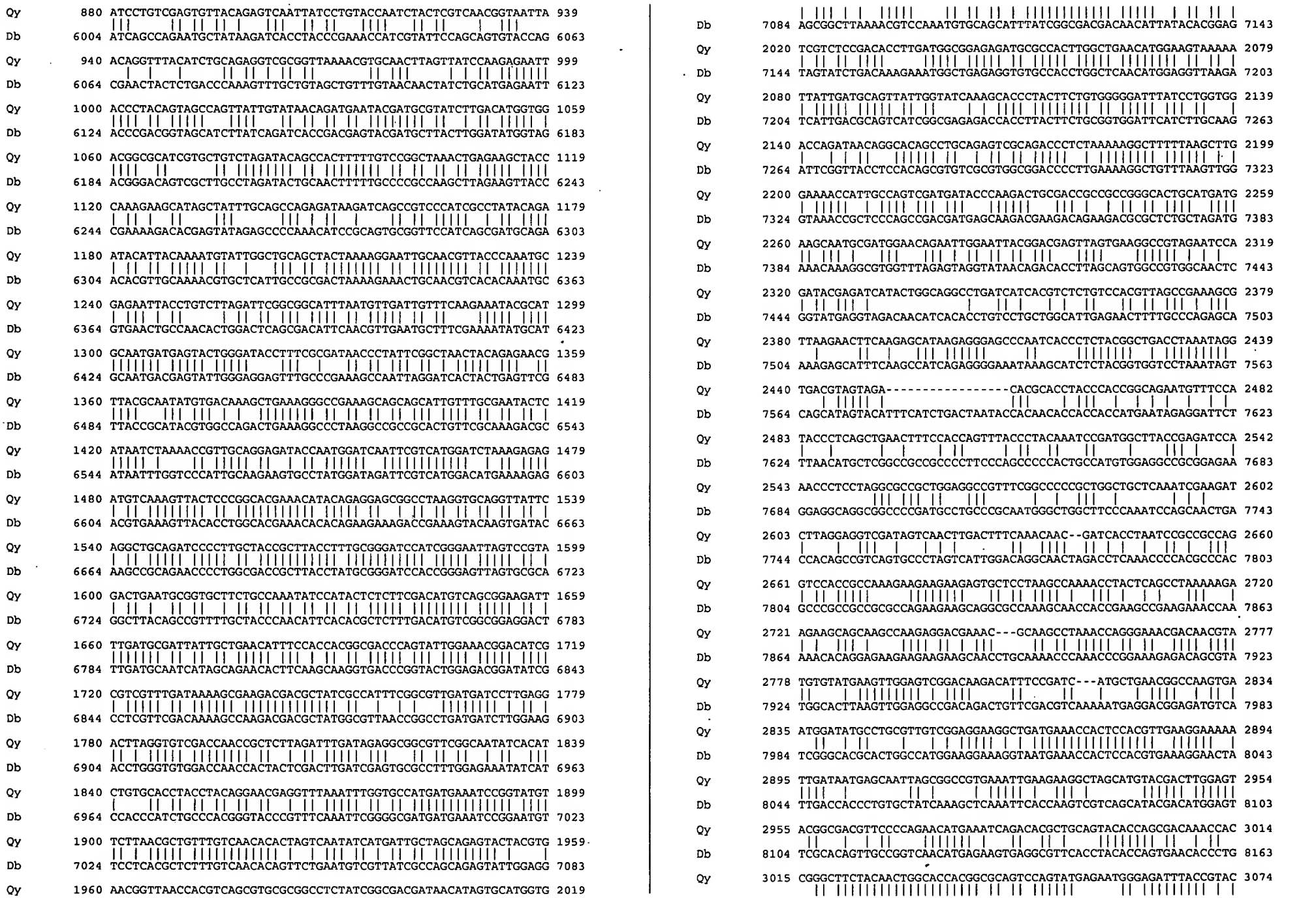

Qy 2698 AAAACCTACTCAGCCTAAAAAGAAGAAGCAGCAAGCCAAGAGGACGAAACGCAAGCCTAA 2757
Db 7690 AAAACCTACTCAGCCTAAAAAGAAGAAGCAGCAAGCCAACAAGACGAAACGCAAGCCTAA 7749
Qy 2758 ACCAGGGAAACGACAACGTATGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATCAT 2817
Db 7750 ACCAGGGAAACGACAGCGTATGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATCAT 7809
Qy 2818 GCTGAACGGCCAAGTGAATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACT 2877
Db 7810 GTTGAACGGCCAAGTGAATGGATACGCTTGCGTTGTCGGAGGAAGGCTGATGAAACCACT 7869
Qy 2878 CCACGTTGAAGGAAAAATTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAG 2937
Db 7870 CCACGTTGAAGGAAAAATCGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAG 7929
Qy 2938 CATGTACGACTTGGAGTACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTA 2997
Db 7930 CATGTACGACCTGGAGTATGGCGACGTTCCCCAGAATATGAAATCAGACACGCTGCAGTA 7989
Qy 2998 CACCAGCGACAAACCACCGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAA 3057
Db 7990 CACCAGCGACAAACCACCGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAA 8049
Qy 3058 TGGGAGATTTACCGTACCGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCT 3117
Db 8050 TGGGAGATTCACCGTACCGCGAGGAGTGGGCGGGAAAGGCGACAGTGGAAGACCGATCCT 8109
Qy 3118 GGACAACAGAGGCAGAGTTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTAC 3177
Db 8110 GGACAACAGAGGCAGAGTTGTGGCTATTGTTCTAGGAGGTGCAAACGAGGGCACGCGTAC 8169
Qy 3178 GGCGCTTTCAGTGGTCACTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGG 3237
Db 8170 GGCGCTTTCAGTGGTCACTTGGAACCAGAAAGGGGTGACCATCAAGGATACCCCCGAAGG 8229
Qy 3238 TTCTGAACCGTGGTCACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATG 3297
Db 8230 TTCTGAACCGTGGTCACTAGTTACAGCGCTGTGCGTGCTTTCGAATGTCACATTCCCTTG 8289
Qy 3298 CGACAAACCACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGA 3357
Db 8290 CGACAAACCACCCGTGTGCTATTCACTGGCGCCAGAACGAACACTCGACGTGCTCGAGGA 8349
Qy 3358 GAACGTCGACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAATGTCCATCACG 3417
Db 8350 GAACGTCGACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAATGTCCATCACG 8409
Qy 3418 CCGGCCCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTG 3477
Db 8410 CCGGCCCAAACGAAGCATTACCGATGACTTCACGCTGACCAGTCCCTACCTGGGGTTCTG 8469
Qy 3478 CCCGTATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGA 3537
Db 8470 CCCGTATTGCAGACACTCAGCGCCATGTTTTAGCCCAATAAAAATTGAGAACGTGTGGGA 8529
Qy 3538 CGAATCTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGC 3597
Db 8530 CGAATCTGATGATGGGTCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGC 8589
Qy 3598 AGGCACTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAA 3657
Db 8590 AGGCACTGCAGACGTCACCAAGTTCCGGTACATGTCTTACGACCACGACCATGACATCAA 8649
Qy 3658 GGAAGACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCA 3717
Db 8650 GGAAGACAGTATGGAGAAAATAGCTATTAGTACATCTGGACCATGCCGTCGTCTTGGCCA 8709
Qy 3718 CAAAGGGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCAC 3777
Db 8710 CAAAGGGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCAC 8769
Qy 3778 GAGCGGAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGG 3837

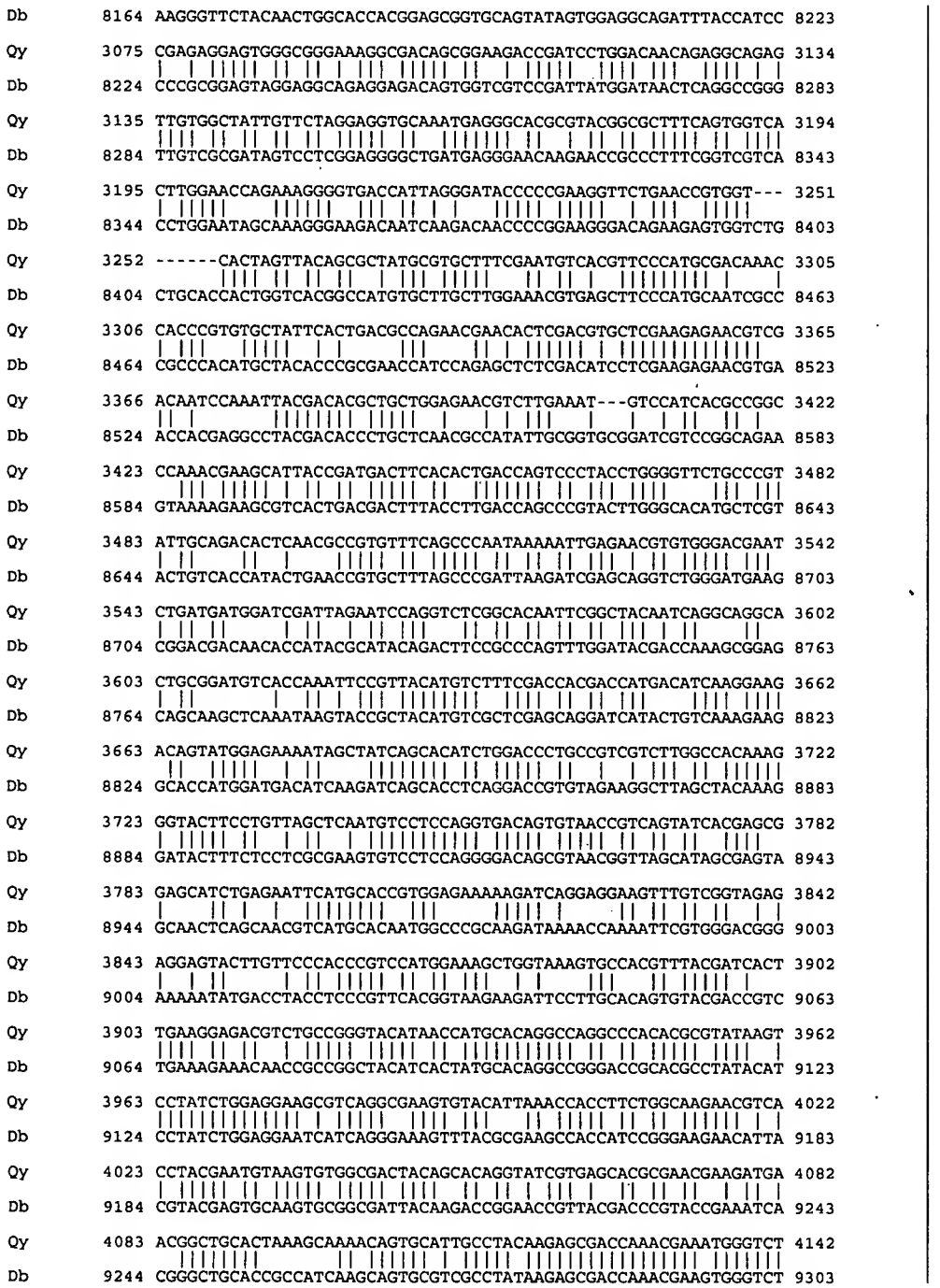

Db 8770 GAGCGGAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGG 8829
Qy 3838 TAGAGAGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGA 3897
Db 8830 TAGAGAGGAGTACTTGTTCCCACCTGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGA 8889
Qy 3898 TCACTTGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTA 3957
Db 8890 TCACTTGAAGGAGACGTCTGCCGGATACATAACCATGCACAGGCCAGGCCCACACGCGTA 8949
Qy 3958 TAAGTCCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAA 4017
Db 8950 TAAGTCCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAA 9009
Qy 4018 CGTCACCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAA 4077
Db 9010 CGTCACCTACGAATGTAAGTGTGGTGACTACAGCACAGGTATTGTGAGCACGCGAACGAA 9069
Qy 4078 GATGAACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATG 4137
Db 9070 GATGAACGGCTGCACTAAAGCAAAACAATGCATTGCCTACAAGCGCGACCAAACGAAATG 9129
Qy 4138 GGTCTTCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCA 4197
Db 9130 GGTCTTCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAACTGCA 9189
Qy 4198 CATTCCATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGT 4257
Db 9190 CATTCCATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGT 9249
Qy 4258 CACGAAGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGAC 4317
Db 9250 CACGAAGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAACGCGACCAACATTGCTGAC 9309
Qy 4318 AACGAGAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATC 4377
Db 9310 AACGAGAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGACTACATC 9369
Qy 4378 CAGGAATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGT 4437
Db 9370 CAGGAATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGCAACCATGAACCAGT 9429
Qy 4438 CAGAGTCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCAT 4497
Db 9430 CAGAGTCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCAT 9489
Qy 4498 CATCCACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCT 4557
Db 9490 CATCCACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCT 9549
Qy 4558 TGCTATCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTG 4617
Db 9550 TGCTATCCTGGTAGGCACTGCATCGTCAGCAGCTTGTATCGCCAAAGCAAGAAGAGACTG 9609
Qy 4618 CCTGACGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTG 4677
Db 9610 CCTGACGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCAGTTTTGTG 9669
Qy 4678 CTGCATTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAA 4737
Db 9670 CTGTATTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAA 9729
Qy 4738 CAACCAACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTT 4797
Db 9730 CAACCAACCGTTTCTCTGGGCACAGTTGTGCATCCCTCTGGCAGCGCTTGTTATTCTGTT 9789
Qy 4798 CCGCTGCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGT 4857
Db 9790 CCGCTGCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGT 9849
Qy 4858 AGACGCCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTT 4917
Db 9850 AGACGCCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTT 9909
Qy 4918 GGTCGAACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATT 4977
Db 9910 GGTCGAACGTGCAGGTTACGCGCCACTTAATCTGGAGATTACGGTCGTCTCATCGGAATT 9969
Qy 4978 AACACCTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACC 5037
Db 9970 AACACCCTCAACTAACAAGGAGTACGTGACCTGCAAATTTCACACAGTCGTTCCTTCACC 10029
Qy 5038 ACAAGTTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATG 5097
Db 10030 ACAAGTTAAATGCTGCGGGTCCCTCGAGTGTAAGGCATCCTCAAAAGCGGATTACACATG 10089
Qy 5098 CCGCGTTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAG 5157
Db 10090 CCGCGTTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAGTGCTTCTGTGACAG 10149
Qy 5158 TGAGAACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCA 5217
Db 10150 TGAGAACACACAACTGAGTGAGGCATACGTCGAGTTCGCTCCAGACTGCACTATAGATCA 10209
Qy 5218 CGCAGTCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGG 5277
Db 10210 TGCAGTCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGG 10269
Qy 5278 CAACACCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGA 5337
Db 10270 CAATACCACAGCGCGCCTGGATACATTCGTCAACGGCGTCACACCAGGTTCCTCACGGGA 10329
Qy 5338 CCTGAAGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGT 5397
Db 10330 CCTGAAGGTCATAGCAGGGCCGATATCAGCAGCTTTTTCACCCTTTGACCATAAGGTCGT 10389
Qy 5398 CATCAGAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGG 5457
Db 10390 CATTAGAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAACCCAGG 10449
Qy 5458 AGCGTTCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGA 5517
Db 10450 AGCGTTCGGCGATATTCAAGCATCCTCTCTTGATGCCACAGACATAGTAGCCCGCACCGA 10509
Qy 5518 CATACGGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATC 5577
Db 10510 CATACGGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATC 10569
Qy 5578 AGGGTATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGG 5637
Db 10570 AGGGTATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTCGG 10629
Qy 5638 ATGTAAAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTAT 5697
Db 10630 ATGTAAAATTGAAGTGGAGCCTCTGCGAGCGACTAACTGTGCTTATGGGCACATCCCTAT 10689
Qy 5698 CTCGATTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGA 5757
Db 10690 CTCGATTGACATCCCTGATGCAGCTTTTGTGAGATCATCTGAATCACCAACAATTTTAGA 10749
Qy 5758 AGTTAGCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATT 5817
Db 10750 AGTCAGCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCGCTAACACT 10809
Qy 5818 ACAGTACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGT 5877
Db 10810 ACAGTACAAAGCTAACAGAGAGGGACATTGTCCAGTTCACTCCCACTCCACTACAGCTGT 10869
Qy 5878 TTTGAAGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCAC 5937
Db 10870 TTTGAAGGAAGCGACCACACATGTGACTGCCACAG---CCATAACACTACATTTTAGCAC 10926
Qy 5938 ATCGAGCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGC 5997
Db 10927 ATCGAGCCCACAAGCAAATTTCATAGTTTCGCTATGGCGCAAGAAGACCACCTGCAATGC 10986

```
Qy       5998 TGAATGTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATT 6057
             ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      10987 TGAATGTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAGGTCGACCAAGAATT 11046

Qy       6058 CCAGGCGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATC 6117
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11047 CCAGGCGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATC 11106

Qy       6118 ATCCCTCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACG 6177
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11107 ATCCCTCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACG 11166

Qy       6178 TAGATGACTGAGCGCGGACACTGACATAGCGGT-AAAACTCGATGTACTTCCGAGGAAGC 6236
             |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db      11167 TAGATGACTGAGCGCGGACACTGACATAGCGGTAAAAACTCGATGTACTTCCGAGGAAGC 11226

Qy       6237 GTGGTGCATAATGCCACGCGCCGCTTGACACTAAAACTCGATGTATTTCCGAGGAAGCAC 6296
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11227 GTGGTGCATAATGCCACGCGCCGCTTGACACTAAAACTCGATGTATTTCCGAGGAAGCAC 11286

Qy       6297 AGTGCATAATGCTGTGCAGTGTCACATTAATCGTATATCACACTACATATTAACAACACT 6356
             ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      11287 AGTGCATAATGCTGTGCAGTGTCACATTAATCGTATATTACACTACATATTAACAACACT 11346

Qy       6357 ATATCACTTTTATGAGACTCACTATGGGTCTCTAATATACACTACACATATTTTACTTAA 6416
             |||||||||||||||||||||||||| ||||||| |||||||||| |||||| ||||
Db      11347 ATATCACTTTTATGAGACTCACTATGGGTTTCTAATACACACTACACACATTTTATTTAA 11406

Qy       6417 AAACACTATACACACTTTATAAATTCTTTTATAATTTTTCTTTTGTTTTTATTTTGTTTT 6476
             |||||||| ||||||||||||||||||||||||||||||||||||||  |||  |||
Db      11407 AAACACTACACACACTTTATAAATTCTTTTATAATTTTTCTTTTGTTTTTATTTTGTTT 11466

Qy       6477 TAAAATTT 6484
             | ||| ||
Db      11467 TTAAAATT 11474
```

RESULT 2
AAV33475
ID AAV33475 standard; cDNA; 11663 BP.
XX
AC AAV33475;
XX
DT 10-DEC-1998 (first entry)
XX
DE South African Arbovirus strain No. 86 cDNA clone pS55 sequence.
XX
KW Bone marrow cell; alphavirus; South African Arbovirus strain No. 86;
KW S.A.AR86; reverse transcriptase-PCR; nsP1; nsP2; nsP3; nsP4; capsid;
KW E3; E2; 6K; E1; growth hormone; growth factor; interleukin; cytokine;
KW chemokine; enzyme; ribozyme; antisense oligonucleotide; ss.
XX
OS South african arbovirus.
XX
FH Key   Location/Qualifiers
FT variation   215
FT   /*tag= a
FT   /note= "This nucleotide varies from A to G with respect
FT   to the S.A.AR86 cDNA sequence given in AAV33473"
FT variation   3863
FT   /*tag= b
FT   /note= "This nucleotide varies from C to G with respect
FT   to the S.A.AR86 cDNA sequence given in AAV33473"
FT variation   5984
FT   /*tag= c
FT   /note= "This nucleotide varies from A to G with respect
FT   to the S.A.AR86 cDNA sequence given in AAV33473"
FT variation   9113
FT   /*tag= d
FT   /note= "This nucleotide varies from T to C with respect
FT   to the S.A.AR86 cDNA sequence given in AAV33473"
XX
PN WO9836779-A2.
XX
PD 27-AUG-1998.
XX
PF 18-FEB-1998; 98WO-US02945.
XX
PR 19-FEB-1997; 97US-0801263.
XX
PA (UYNC-) UNIV NORTH CAROLINA.
XX
PI Davis NL, Johnston RE, Simpson DA;
XX
DR WPI; 1998-495361/42.
DR P-PSDB; AAW70471, AAW70472.
XX
PT Expression of heterologous RNA in bone marrow cells - using a
PT recombinant alphavirus comprising a promoter operable in bone marrow
PT cells operably associated with heterologous RNA.
XX
PS Example 5; Fig 5A-5B; 68pp; English.
XX
CC The present sequence represents the South African Arbovirus strain
CC No. 86 (S.A.AR86) cDNA clone pS55 sequence used in the method of
CC the invention. The invention provides a method for introducing and
CC expressing heterologous RNA in bone marrow cells using alphavirus
CC vectors. S.A.AR86 is an alphavirus whose cDNA sequence (AAV33474) was
CC determined from uncloned reverse transcriptase-PCR reaction fragments
CC amplified from the virion RNA. The S.A.AR86 cDNA clone pS55 sequence
CC has four nucleotide variations when compared with the S.A.AR86 cDNA
CC sequence derived from genomic RNA. The variations, however, do not
CC alter the the sequence of the protein encoded with respect to the
CC S.A.SR86 genomic RNA. The S.A.AR86 cDNA encodes nonstructural
CC and structural polyproteins. The nonstructural polyprotein is,
CC presumably, post-translationally modified into four different protein
CC products, namely nsP1 (AAW70460), nsP2 (AAW70461), nsP3 (AAW70466), nsP4
CC (AAW70467) proteins. The structural polyprotein is, presumably,
CC post-translationally modified into five different protein products,
CC namely capsid (AAW70468), E3 (AAW70469), E2 (AAW70470), 6K (AAW70471),
CC E1 (AAW70472) proteins. The S.A.AR86 cDNA was used in the method of the
CC invention. The inventors claim the transformed bone marrow cells are
CC useful for expressing a protein or peptide suitable for protecting the
CC subject against a disease such as a microbial, bacterial, protozoal,
CC parasitic or viral disease. The transformed bone marrow cells are also
CC claimed to be useful for expressing proteins and peptides such as
CC hormones, growth hormones, growth factors, interleukins, cytokines,
CC chemokines, enzymes, ribozymes or antisense oligonucleotides.
XX
SQ Sequence 11663 BP; 3289 A; 3105 C; 2910 G; 2359 T; 0 other;

```
  Query Match            37.0%;  Score 2397.6;  DB 19;  Length 11663;
  Best Local Similarity  64.6%;  Pred. No. 0;
  Matches 3752;  Conservative    0;  Mismatches 1984;  Indels   74;  Gaps    9;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
              |||| |     |  |  || | ||| || || ||  |||| |||||||  || ||| | |
Db       5707 ACTGTCTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5766

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
              || |||| || || |  ||     | ||   |||||| |  ||||| ||    || |  |
Db       5767 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTTCTGG 5826

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
              | |    |||||||||   |||||       | |||| ||  | |  |    |     |
Db       5827 AAAGAATCTACGCCCCGGTGCTCGACACGTCGAAAGAGGAACAGCTCAAACTCAGGTACC 5886

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
              |  |     || | |||  || | |||||||||| || || ||||||||||||||   ||
Db       5887 AGATGATGCCCACCGAAGCCAACAAAAGCAGGTACCAGTCTCGAAAAGTAGAAAACCAGA 5946

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
              |||| || ||  | |||||||| |||| ||  |  |    |||   ||  |   ||
Db       5947 AAGCCATAACCACTGAGCGACTGCTTTCAGGGCTACGGCTGTATAACTCTGCCACAG--- 6003
```



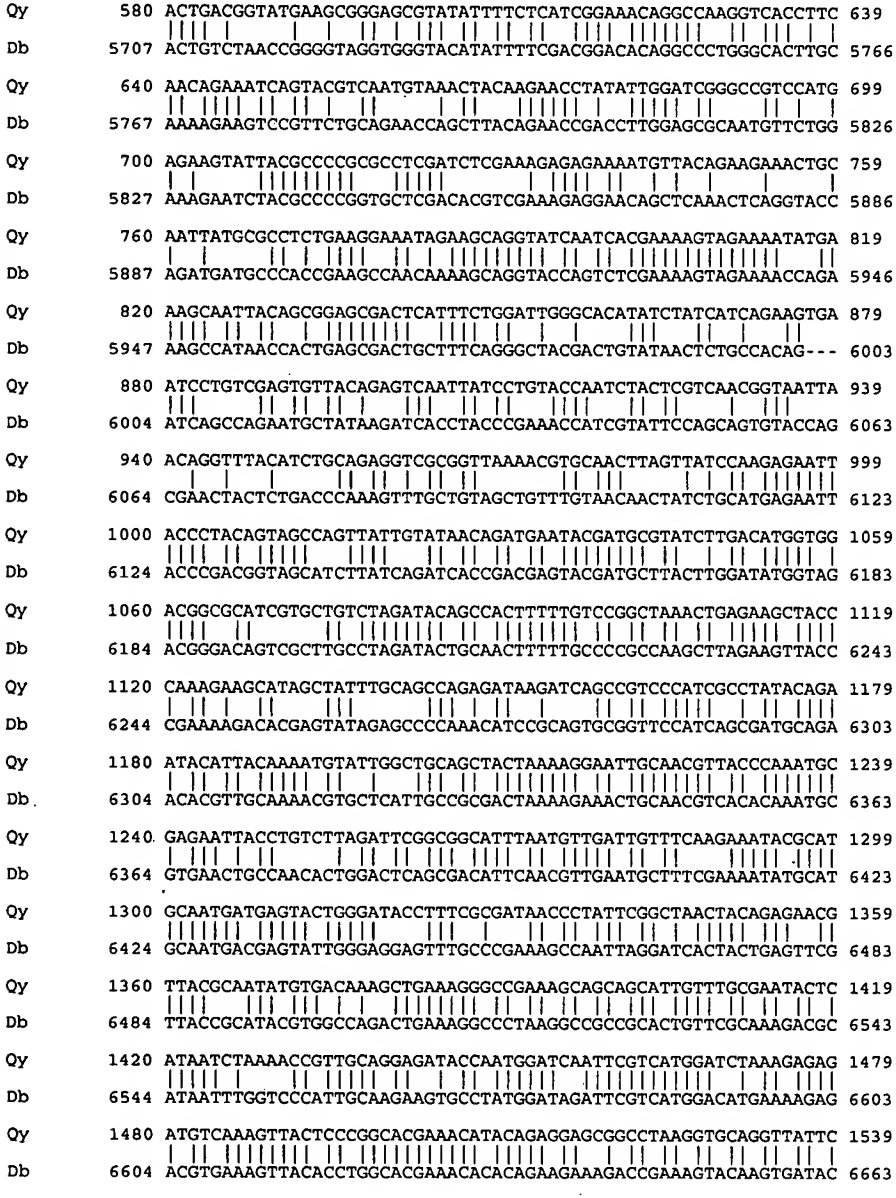

Db 8164 AAGGGTTCTACAACTGGCACCACGGAGCGGTGCAGTATAGTGGAGGCAGATTTACCATCC 8223
Qy 3075 CGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGACAACAGAGGCAGAG 3134
Db 8224 CCCGCGGAGTAGGAGGCAGAGGAGACAGTGGTCGTCCGATTATGGATAACTCAGGCCGGG 8283
Qy 3135 TTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGCGCTTTCAGTGGTCA 3194
Db 8284 TTGTCGCGATAGTCCTCGGAGGGGCTGATGAGGGAACAAGAACCGCCCTTTCGGTCGTCA 8343
Qy 3195 CTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTCTGAACCGTGGT--- 3251
Db 8344 CCTGGAATAGCAAAGGGAAGACAATCAAGACAACCCCGGAAGGGACAGAAGAGTGGTCTG 8403
Qy 3252 ------CACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGACAAAC 3305
Db 8404 CTGCACCACTGGTCACGGCCATGTGCTTGCTTGGAAACGTGAGCTTCCCATGCAATCGCC 8463
Qy 3306 CACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAACGTCG 3365
Db 8464 CGCCCACATGCTACACCCGCGAACCATCCAGAGCTCTCGACATCCTCGAAGAGAACGTGA 8523
Qy 3366 ACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAAT---GTCCATCACGCCGGC 3422
Db 8524 ACCACGAGGCCTACGACACCCTGCTCAACGCCATATTGCGGTGCGGATCGTCCGGCAGAA 8583
Qy 3423 CCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCCCGT 3482
Db 8584 GTAAAAGAAGCGTCACTGACGACTTTACCTTGACCAGCCCGTACTTGGGCACATGCTCGT 8643
Qy 3483 ATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACGAAT 3542
Db 8644 ACTGTCACCATACTGAACCGTGCTTTAGCCCGATTAAGATCGAGCAGGTCTGGGATGAAG 8703
Qy 3543 CTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAGGCA 3602
Db 8704 CGGACGACAACACCATACGCATACAGACTTCCGCCCAGTTTGGATACGACCAAAGCGGAG 8763
Qy 3603 CTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAAGGAAG 3662
Db 8764 CAGCAAGCTCAAATAAGTACCGCTACATGTCGCTCGAGCAGGATCATACTGTCAAAGAAG 8823
Qy 3663 ACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCACAAAG 3722
Db 8824 GCACCATGGATGACATCAAGATCAGCACCTCAGGACCGTGTAGAAGGCTTAGCTACAAAG 8883
Qy 3723 GGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCACGAGCG 3782
Db 8884 GATACTTTCTCCTCGCGAAGTGTCCTCCAGGGGACAGCGTAACGGTTAGCATAGCGAGTA 8943
Qy 3783 GAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTAGAG 3842
Db 8944 GCAACTCAGCAACGTCATGCACAATGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGG 9003
Qy 3843 AGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATCACT 3902
Db 9004 AAAAATATGACCTACCTCCCGTTCACGGTAAGAAGATTCCTTGCACAGTGTACGACCGTC 9063
Qy 3903 TGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAAGT 3962
Db 9064 TGAAAGAAACAACCGCCGGCTACATCACTATGCACAGGCCGGGACCGCACGCCTATACAT 9123
Qy 3963 CCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGTCA 4022
Db 9124 CCTATCTGGAGGAATCATCAGGGAAAGTTTACGCGAAGCCACCATCCGGGAAGAACATTA 9183
Qy 4023 CCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGATGA 4082
Db 9184 CGTACGAGTGCAAGTGCGGCGATTACAAGACCGGAACCGTTACGACCCGTACCGAAATCA 9243
Qy 4083 ACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGTCT 4142
Db 9244 CGGGCTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAACGAAGTGGGTCT 9303
Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9304 TCAACTCGCCGGACTCGATCAGACACGCCGACCACACGGCCCAAGGGAAATTGCATTTGC 9363
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9364 CTTTCAAGCTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAACGTAGTAC 9423
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9424 ACGGCTTTAAACACATCAGCCTCCAATTAGACACAGACCATCTGACATTGCTCACCACCA 9483
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9484 GGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCATCGGAAACACGGTTAGAA 9543
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9544 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGCAATCACGAACCAGTAAGGG 9603
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db 9604 TCTATGCCCAAGAGTCTGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9663
Qy 4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db 9664 ATTACTATCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTGCTGTGGCGA 9723
Qy 4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db 9724 TGATGATTGGCGTAACTGTTGCAGCATTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9783
Qy 4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db 9784 CGCCATATGCCCTGGCCCCAAATGCCGTGATTCCAACTTCGCTGGCACTTTTGTGCTGTG 9843
Qy 4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db 9844 TTAGGTCGGCTAATGCTGAAACATTCACCGAGACCATGAGTTACTTATGGTCGAACAGCC 9903
Qy 4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
Db 9904 AGCCGTTCTTCTGGGTCCAGCTGTGTATACCTCTGGCCGCTGTCGTCGTTCTAATGCGCT 9963
Qy 4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
Db 9964 GTTGCTCATGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10023
Qy 4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
Db 10024 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10083
Qy 4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db 10084 AAAGGGCAGGGTACGCCCCGCTCAATTTGGAGATTACTGTCATGTCCTCGGAGGTTTTGC 10143
Qy 4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db 10144 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCTAAAG 10203
Qy 5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db 10204 TCAGATGCTGCGGCTCCTTGGAATGTCAGCCCGCCGCTCACGCAGACTATACCTGCAAGG 10263
Qy 5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db 10264 TCTTTGGAGGGGTGTACCCCTTCATGTGGGGAGGAGCACAATGTTTTTGCGACAGTGAGA 10323
Qy 5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db 10324 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGTAGATTGCGCGACTGACCACGCGC 10383

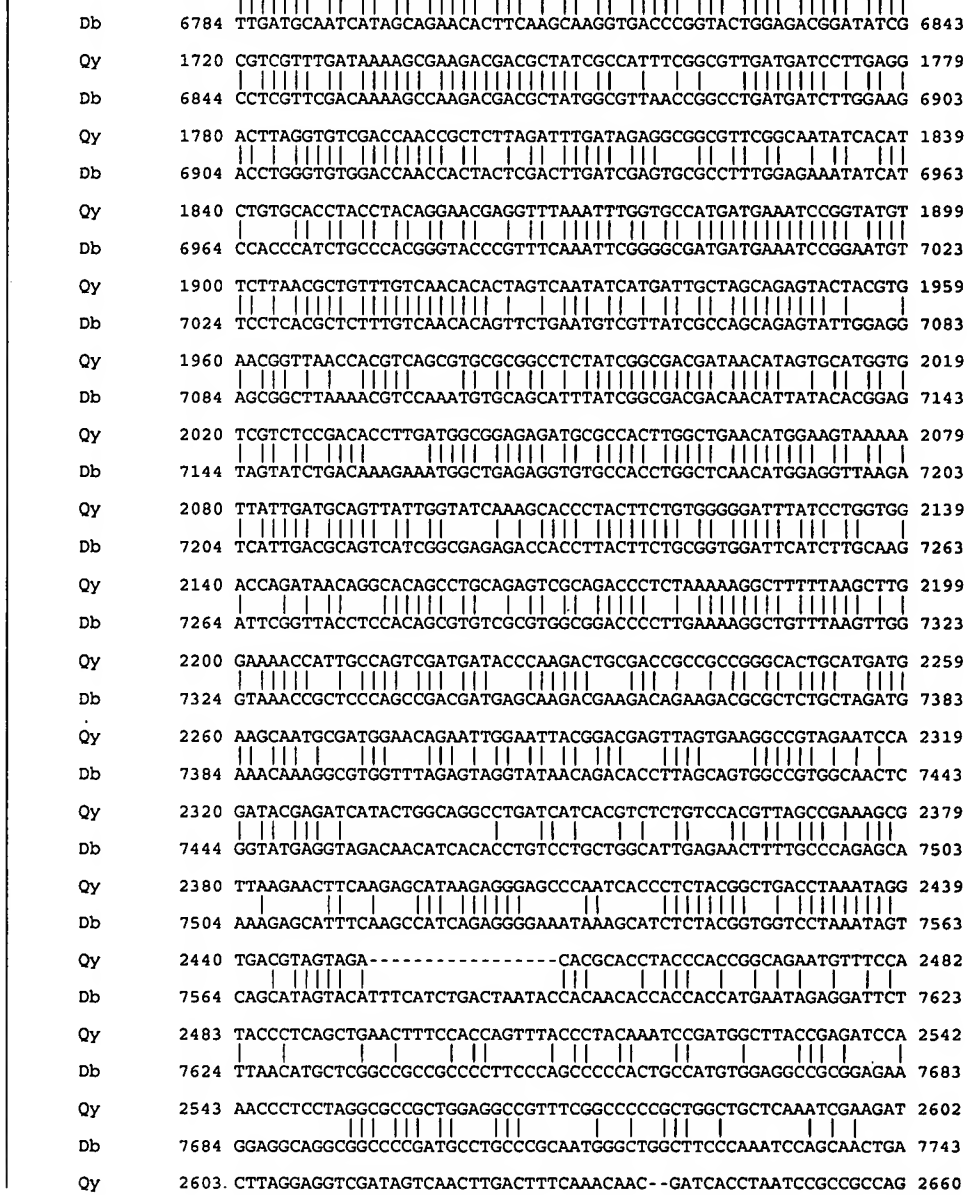
```
Qy       5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db      10384 AGGCGATTAAGGTGCATACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10443

Qy       5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db      10444 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACCTGA 10503

Qy       5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db      10504 AAGTCATAGCTGGACCAATTTCAGCATTGTTTACACCATTCGATCACAAGGTCGTTATCA 10563

Qy       5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db      10564 ATCGCGGCCTGGTGTACAACTATGACTTTCCGGAATACGGAGCGATGAAACCAGGAGCGT 10623

Qy       5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db      10624 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAAGACCTCATCGCCAGCACAGACATTA 10683

Qy       5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db      10684 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCTGGAT 10743

Qy       5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db      10744 TCGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCCCCTTTTGGGTGCA 10803

Qy       5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db      10804 AGATTGCAGTCAATCCGCTTCGAGCGGTGGACTGCTCATACGGGAACATTCCCATTTCTA 10863

Qy       5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db      10864 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10923

Qy       5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db      10924 AATGTGATGTCAGTGAGTGCACTTATTCAGCGGACTTCGGAGGGATGGCTACCCTGCAGT 10983

Qy       5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db      10984 ATGTATCCGACCGCGAAGGACAATGCCCTGTACATTCGCATTCGAGCACAGCAACCCTCC 11043

Qy       5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db      11044 AAGAGTCGACAGTTCATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTCAGCACCGCGA 11103

Qy       5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db      11104 GCCCACAGGCGAACTTCATTGTATCGCTGTGTGGTAAGAAGACAACATGCAATGCAGAAT 11163

Qy       6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db      11164 GCAAACCACCAGCTGATCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTCCAAG 11223

Qy       6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db      11224 CCGCCATCTCAAAAACTTCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11283

Qy       6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
Db      11284 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11343

Qy       6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
Db      11344 GACCGCTACGCCCCAATGACCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11403

Qy       6242 GCATAATGCCACGCGCCGCT--------------------------------TGACACT 6268
Db      11404 GCATAATGCATCAGGCTGGTATATTAGATCCCCGCTTACCGCGGGCAATATAGCAACACC 11463

Qy       6269 AAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGT 6318
                ||||||||| |||||||||||||||| |||||||||||||| |||||||
Db      11464 AAAACTCGACGTATTTCCGAGGAAGCGCAGTGCATAATGCTGCGCAGTGT 11513
```

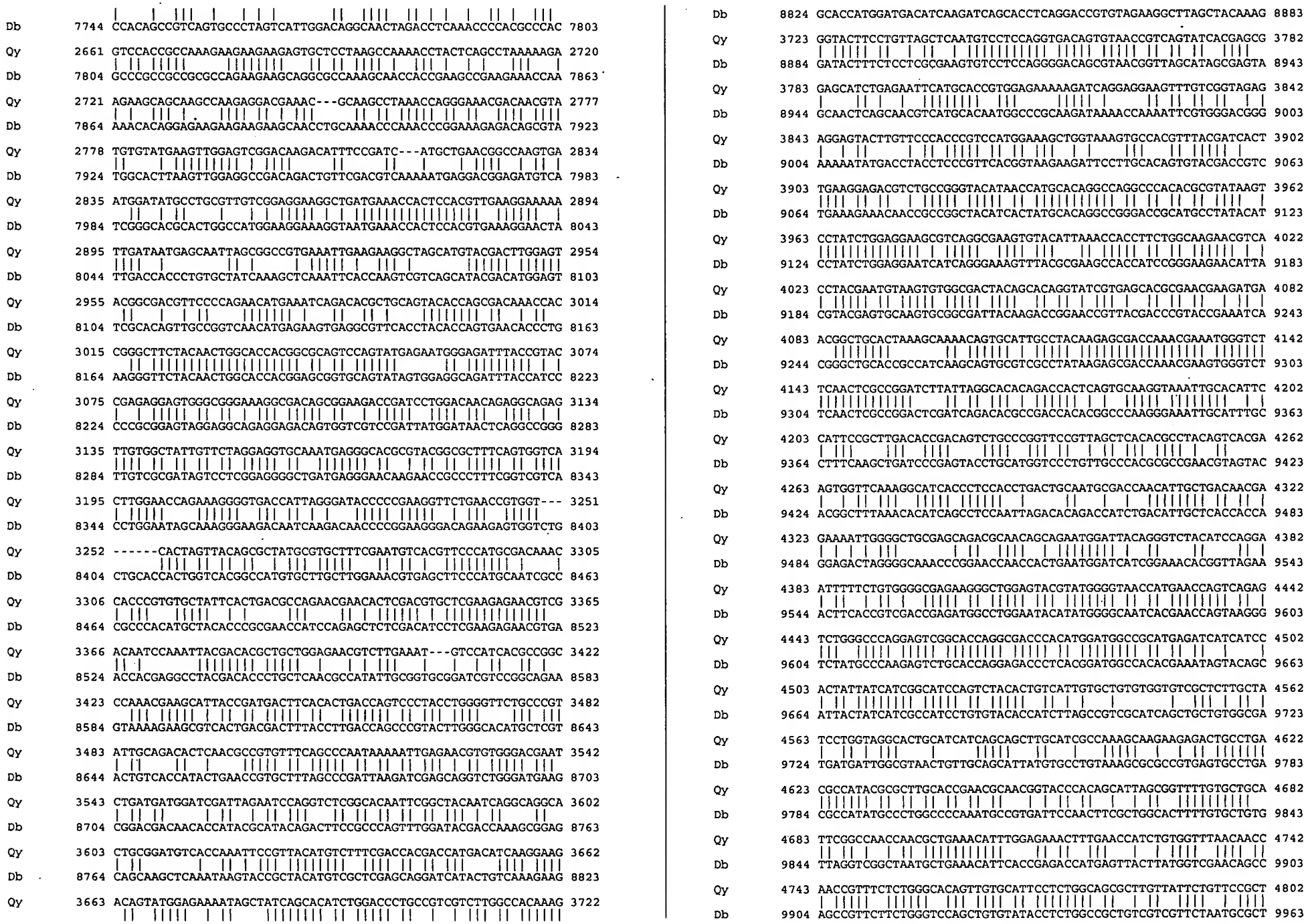
RESULT 3
AAV33473
ID AAV33473 standard; cDNA; 11663 BP.
XX
AC AAV33473;
XX
DT 10-DEC-1998 (first entry)
XX
DE South African Arbovirus strain No. 86 cDNA sequence.
XX
KW Bone marrow cell; alphavirus; South African Arbovirus strain No. 86;
KW S.A.AR86; reverse transcriptase-PCR; nsP1; nsP2; nsP3; nsP4; capsid;
KW E3; E2; 6K; E1; growth hormone; growth factor; interleukin; cytokine;
KW chemokine; enzyme; ribozyme; antisense oligonucleotide; ss.
XX
OS South african arbovirus
XX
FH Key Location/Qualifiers
FT 5'UTR 1..59
FT /*tag= a
FT CDS 60..7662
FT /*tag= b
FT /product= "Non-structural polyprotein"
FT CDS 7608..11345
FT /*tag= c
FT /product= "Structural polyprotein"
FT 3'UTR 11346..11663
FT /*tag= d
XX
PN WO9836779-A2.
XX
PD 27-AUG-1998.
XX
PF 18-FEB-1998; 98WO-US02945.
XX
PR 19-FEB-1997; 97US-0801263.
XX
PA (UYNC-) UNIV NORTH CAROLINA.
XX
PI Davis NL, Johnston RE, Simpson DA;
XX
DR WPI; 1998-495361/42.
DR P-PSDB; AAW70460, AAW70461, AAW70466, AAW70467, AAW70468, AAW70469,
DR AAW70470, AAW70471, AAW70472.
XX
PT Expression of heterologous RNA in bone marrow cells - using a
PT recombinant alphavirus comprising a promoter operable in bone marrow
PT cells operably associated with heterologous RNA.
XX
PS Example 2; Fig 1A-1C; 68pp; English.
XX
CC The invention provides a method for introducing and expressing
CC heterologous RNA in bone marrow cells using alphavirus vectors.
CC The South African Arbovirus strain No. 86 (S.A.AR86) is an
CC alphavirus whose cDNA sequence was determined from uncloned reverse
CC transcriptase-PCR reaction fragments amplified from the virion RNA.
CC The S.A.AR86 cDNA encodes nonstructural and structural polyproteins.
CC The nonstructural polyprotein is, presumably, post-translationally
CC modified into four different protein products, namely nsP1 (AAW70460),
CC nsP2 (AAW70461), nsP3 (AAW70466), nsP4 (AAW70467) proteins. The
CC structural polyprotein is, presumably, post-translationally modified into
CC five different protein products, namely capsid (AAW70468), E3
CC (AAW70469), E2 (AAW70470), 6K (AAW70471), E1 (AAW70472) proteins. The
CC S.A.AR86 cDNA was used in the method of the invention. The inventors
CC claim the transformed bone marrow cells are useful for expressing a
CC protein or peptide suitable for protecting the subject against a disease
CC such as a microbial, bacterial, protozoal, parasitic or viral disease.
CC The transformed bone marrow cells are also claimed to be useful for
CC expressing proteins and peptides such as hormones, growth hormones,

CC growth factors, interleukins, cytokines, chemokines, enzymes, ribozymes
CC or antisense oligonucleotides.
XX
SQ Sequence 11663 BP; 3291 A; 3105 C; 2907 G; 2360 T; 0 other;

Query Match 36.9%; Score 2396; DB 19; Length 11663;
Best Local Similarity 64.6%; Pred. No. 0;
Matches 3751; Conservative 0; Mismatches 1985; Indels 74; Gaps 9;







```
Qy       4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
             | | |||||||||| |||||||||| ||||||| |||| || ||||| ||||||||||
Db       9964 GTTGCTCATGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10023

Qy       4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
             ||| |||||||||||||||| |||||||| ||   ||| ||||||||||  | || |
Db      10024 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10083

Qy       4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
             || | ||||| ||||| || || ||  ||||||| || ||| | || |||||    |
Db      10084 AAAGGGCAGGGTACGCCCCGCTCAATTTGGAGATTACTGTCATGTCCTCGGAGGTTTTGC 10143

Qy       4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
             |||| || ||| | |||||| | |||||||||||  ||| ||  | || || ||  |||
Db      10144 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCTAAAG 10203

Qy       5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
             | | |||||||||| ||| | || ||  || |  || |  | || || || || |||  |
Db      10204 TCAGATGCTGCGGCTCCTTGGAATGTCAGCCCGCCGCTCACGCAGACTATACCTGCAAGG 10263

Qy       5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
             | ||||| || |||||||| |||||||||||||| |||||||| || || ||||||||||
Db      10264 TCTTTGGAGGGGTGTACCCCTTCATGTGGGGAGGAGCACAATGTTTTTGCGACAGTGAGA 10323

Qy       5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
             |||  ||  |||||||||||||||||||| ||   |||  ||| | |  || |||||
Db      10324 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGTAGATTGCGCGACTGACCACGCGC 10383

Qy       5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
               ||  | || || || || || ||  ||||||| || |||||||||| ||||| ||||
Db      10384 AGGCGATTAAGGTGCATACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10443

Qy       5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
             | |||      ||| |||   ||  || || || |||||  |||||  | ||     |||||||
Db      10444 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACCTGA 10503

Qy       5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
             | ||||||||  || ||  || |||||     ||| |||| || || || ||||||||  ||||
Db      10504 AAGTCATAGCTGGACCAATTTCAGCATTGTTTACACCATTCGATCACAAGGTCGTTATCA 10563

Qy       5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
                || || || |||||||| ||||| || || || ||||| ||||||||||||||||
Db      10564 ATCGCGGCCTGGTGTACAACTATGACTTTCCGGAATACGGAGCGATGAAACCAGGAGCGT 10623

Qy       5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
             | || || ||||||||  ||||  |    | |||| |   | ||| |||| |||||
Db      10624 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAAGACCTCATCGCCAGCACAGACATTA 10683

Qy       5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
             |||| || |||||||| | ||||||| | || ||||| ||||| || || | ||| || |
Db      10684 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCTGGAT 10743

Qy       5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
               || |||||||| |||||||||| || || ||||| ||||| || || ||||| || |
Db      10744 TCGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCCCCTTTTGGGTGCA 10803

Qy       5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
             | |||| ||| | || || |||||     ||||  | |||||| |||| || || || |
Db      10804 AGATTGCAGTCAATCCGCTTCGAGCGGTGGACTGCTCATACGGGAACATTCCCATTTCTA 10863

Qy       5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
             ||||||||||  | || || ||| | ||  |||||||  |||||    | | |  ||| |
Db      10864 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10923

Qy       5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
               ||    ||     || |||| |||||| || ||||| || ||        ||  | ||||
Db      10924 AATGTGATGTCAGTGAGTGCACTTATTCAGCGGACTTCGGAGGGATGGCTACCCTGCAGT 10983

Qy       5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
             |   | | ||| | || ||||| || || || || || || || |  |||||     |
Db      10984 ATGTATCCGACCGCGAAGGACAATGCCCTGTACATTCGCATTCGAGCACAGCAACCCTCC 11043

Qy       5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
             | ||  ||||    |||||     |    |||      | ||| |||| || |||||  |||
Db      11044 AAGAGTCGACAGTTCATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTCAGCACCGCGA 11103

Qy       5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
             ||||||| || || || || || ||||| || || |||||| | || |||||||| ||||
Db      11104 GCCCACAGGCGAACTTCATTGTATCGCTGTGTGGTAAGAAGACAACATGCAATGCAGAAT 11163

Qy       6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
             | |||||||| || || || ||  | |    || || |||   |||||||||||||| |
Db      11164 GCAAACCACCAGCTGATCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTCCAAG 11223

Qy       6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
             | || | || ||||| || ||||  |||||| |||| || || || || || || || |
Db      11224 CCGCCATCTCAAAAACTTCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11283

Qy       6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
             | | || ||||||||||  | || |  ||||||   ||||| |  | ||||| ||||
Db      11284 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11343

Qy       6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
             ||| |        || ||| |  | | | ||||||||||||||||||||||||  |  ||
Db      11344 GACCGCTACGCCCCAATGACCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11403

Qy       6242 GCATAATGCCACGCGCCGCT--------------------------------TGACACT 6268
             |||||||||  |  || | |                                ||||
Db      11404 GCATAATGCATCAGGCTGGTATATTAGATCCCCGCTTACCGCGGGCAATATAGCAACACC 11463

Qy       6269 AAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGT 6318
             ||||||||| |||||||||||||| |||||||||||||| |||||||
Db      11464 AAAACTCGACGTATTTCCGAGGAAGCGCAGTGCATAATGCTGCGCAGTGT 11513
```

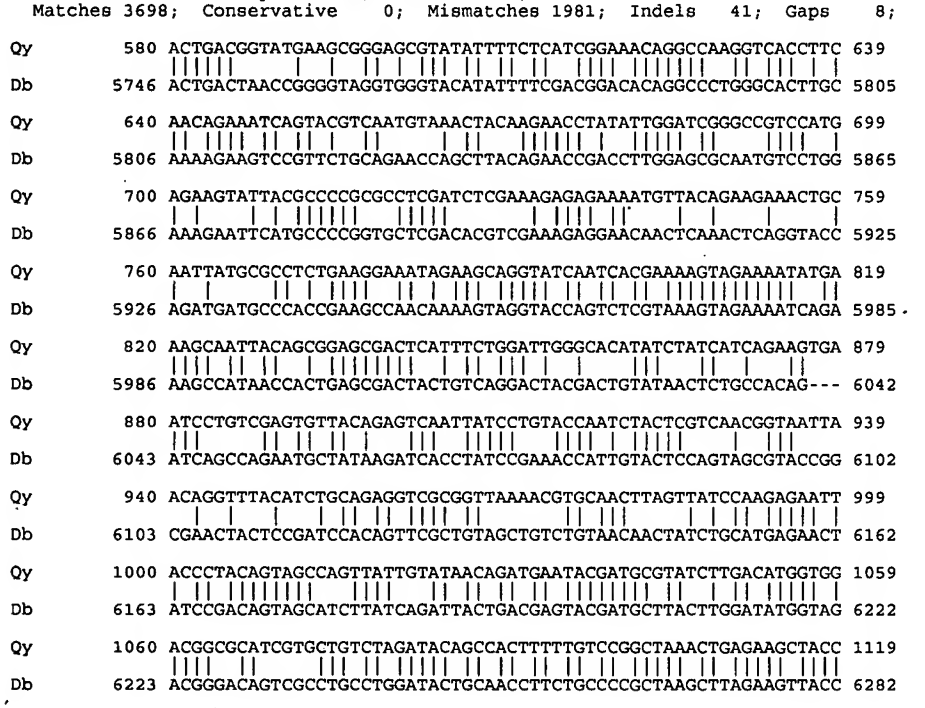
RESULT 4
AAV33476
ID AAV33476 standard; cDNA; 11703 BP.
XX
AC AAV33476;
XX
DT 10-DEC-1998 (first entry)
XX
DE Sindbis virus clone TR339 cDNA sequence.
XX
KW Bone marrow cell; alphavirus; Sindbis virus; cytokine; capsid;
KW reverse transcriptase-PCR; nsP1; nsP2; nsP3; nsP4; E3; E2; 6K; E1;
KW growth hormone; growth factor; interleukin; chemokine; enzyme;
KW ribozyme; antisense oligonucleotide; ss.
XX
OS Sindbis virus.
XX

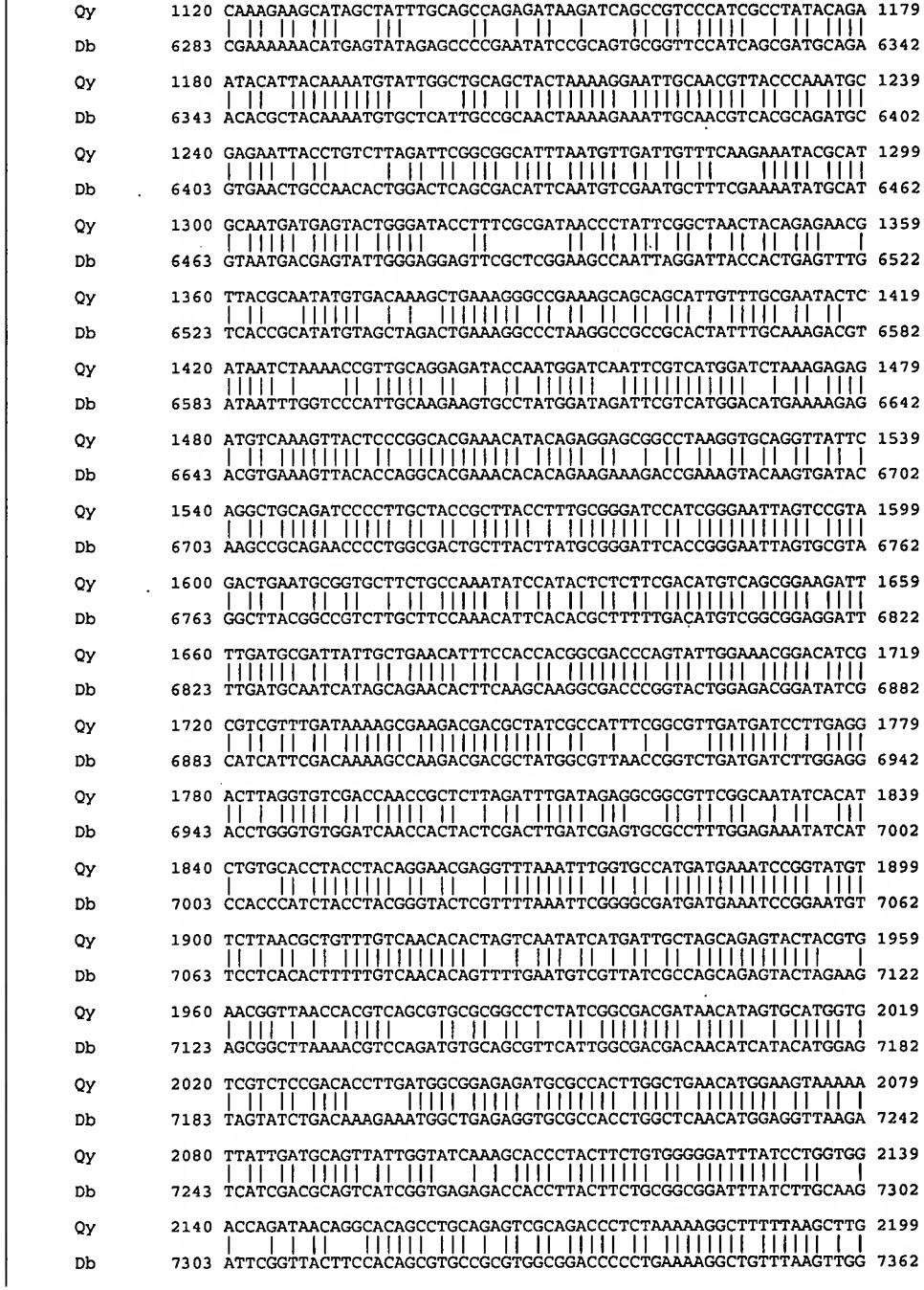
FH	Key	Location/Qualifiers
FT	5'UTR	1..59
FT		/*tag= a
FT	CDS	60..7601
FT		/*tag= b
FT		/product= "Non-structural polyprotein"
FT		/note= "Internal stop codon present at nucleotides 5748-5750"
FT	CDS	7647..11384
FT		/*tag= c
FT		/product= "Structural polyprotein"
FT	3'UTR	11382..11703
FT		/*tag= d

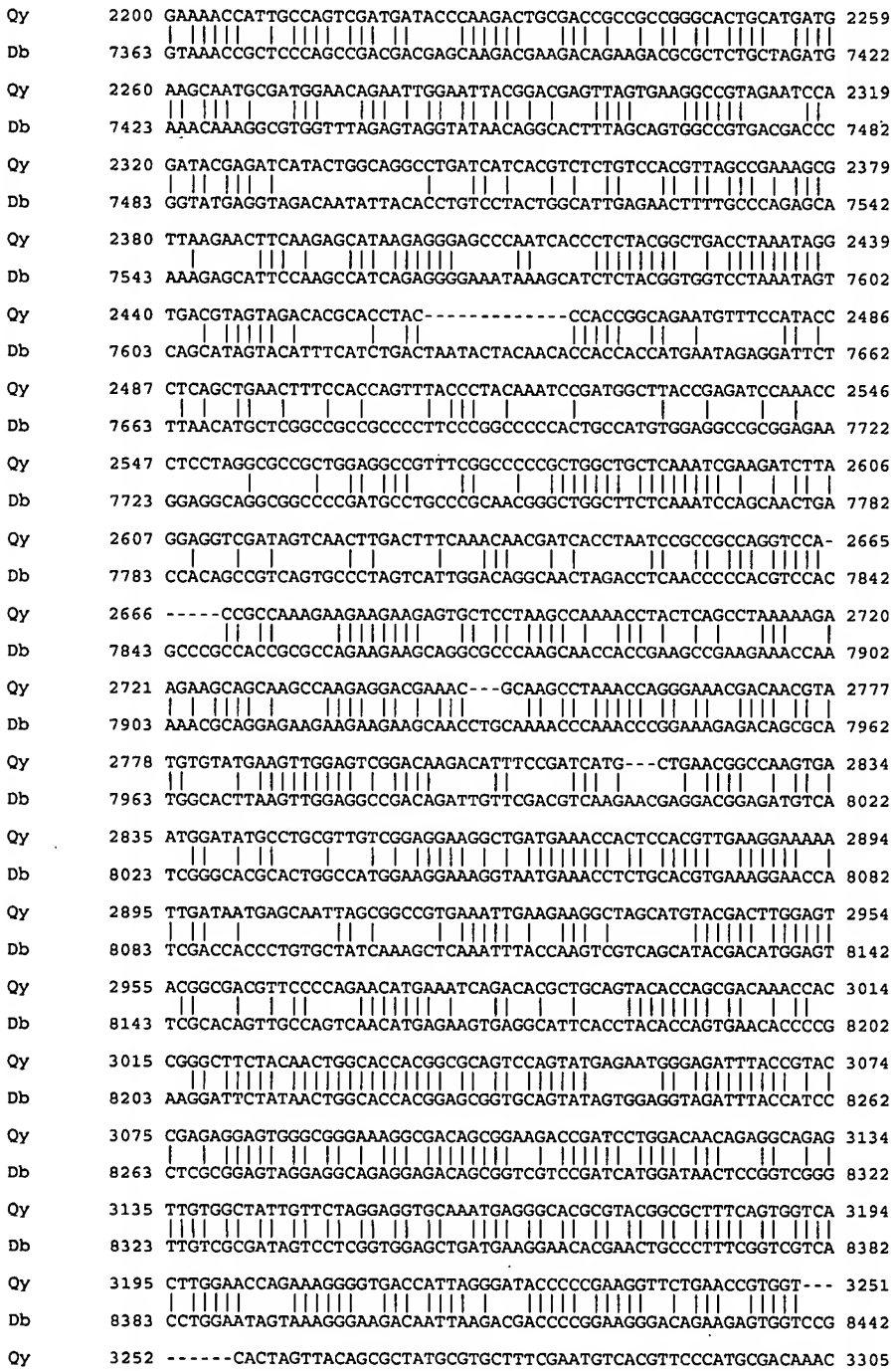
XX
PN WO9836779-A2.
XX
PD 27-AUG-1998.
XX
PF 18-FEB-1998; 98WO-US02945.
XX
PR 19-FEB-1997; 97US-0801263.
XX
PA (UYNC-) UNIV NORTH CAROLINA.

XX
PI Davis NL, Johnston RE, Simpson DA;
XX
DR WPI; 1998-495361/42.
DR P-PSDB; AAW70464, AAW70465, AAW70480, AAW70481, AAW70482, AAW70483,
DR AAW70484, AAW70485, AAW70486.
XX
PT Expression of heterologous RNA in bone marrow cells - using a
PT recombinant alphavirus comprising a promoter operable in bone marrow
PT cells operably associated with heterologous RNA.
XX
PS Example 6; Fig 6A-6B; 68pp; English.
XX
CC The invention provides a method for introducing and expressing
CC heterologous RNA in bone marrow cells using alphavirus vectors.
CC The present sequence represents the Sindbis virus clone TR339 cDNA
CC sequence. The Sindbis virus cDNA encodes nonstructural and structural
CC polyproteins. The nonstructural polyprotein is, presumably,
CC post-translationally modified into four different protein products,
CC namely nsP1 (AAW70464), nsP2 (AAW70465), nsP3 (AAW70480), nsP4
CC (AAW70481) proteins. The structural polyprotein is, presumably, post-
CC translationally modified into five different protein products, namely
CC capsid (AAW70482), E3 (AAW70483), E2 (AAW70484), 6K (AAW70485), E1
CC (AAW70486) proteins. The Sindbis virus cDNA was used in the method of
CC the invention. The inventors claim the transformed bone marrow cells
CC are useful for expressing a protein or peptide suitable for protecting
CC the subject against a disease such as a microbial, bacterial, protozoal,
CC parasitic or viral disease. The transformed bone marrow cells are also
CC claimed to be useful for expressing proteins and peptides such as
CC hormones, growth hormones, growth factors, interleukins, cytokines,
CC chemokines, enzymes, ribozymes or antisense oligonucleotides.
XX
SQ Sequence 11703 BP; 3306 A; 3051 C; 2913 G; 2433 T; 0 other;

Query Match 36.8%; Score 2388.4; DB 19; Length 11703;
Best Local Similarity 64.7%; Pred. No. 0;
Matches 3698; Conservative 0; Mismatches 1981; Indels 41; Gaps 8;






Qy 2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
Db 7363 GTAAACCGCTCCCAGCCGACGACGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 7422
Qy 2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319
Db 7423 AAACAAAGGCGTGGTTTAGAGTAGGTATAACAGGCACTTTAGCAGTGGCCGTGACGACCC 7482
Qy 2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379
Db 7483 GGTATGAGGTAGACAATATTACACCTGTCCTACTGGCATTGAGAACTTTTGCCCAGAGCA 7542
Qy 2380 TTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGG 2439
Db 7543 AAAGAGCATTCCAAGCCATCAGAGGGGAAATAAAGCATCTCTACGGTGGTCCTAAATAGT 7602
Qy 2440 TGACGTAGTAGACACGCACCTAC----------CCACCGGCAGAATGTTTCCATACC 2486
Db 7603 CAGCATAGTACATTTCATCTGACTAATACTACAACACCACCACCATGAATAGAGGATTCT 7662
Qy 2487 CTCAGCTGAACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACC 2546
Db 7663 TTAACATGCTCGGCCGCCGCCCCTTCCCGGCCCCCACTGCCATGTGGAGGCCGCGGAGAA 7722
Qy 2547 CTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTA 2606
Db 7723 GGAGGCAGGCGGCCCCGATGCCTGCCCGCAACGGGCTGGCTTCTCAAATCCAGCAACTGA 7782
Qy 2607 GGAGGTCGATAGTCAACTTGACTTTCAAACAACGATCACCTAATCCGCCGCCAGGTCCA- 2665
Db 7783 CCACAGCCGTCAGTGCCCTAGTCATTGGACAGGCAACTAGACCTCAACCCCCACGTCCAC 7842
Qy 2666 -----CCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGA 2720
Db 7843 GCCCGCCACCGCGCCAGAAGAAGCAGGCGCCCAAGCAACCACCGAAGCCGAAGAAACCAA 7902
Qy 2721 AGAAGCAGCAAGCCAAGAGGACGAAAC---GCAAGCCTAAACCAGGGAAACGACAACGTA 2777
Db 7903 AAACGCAGGAGAAGAAGAAGAAGCAACCTGCAAAACCCAAACCCGGAAAGAGACAGCGCA 7962
Qy 2778 TGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATCATG---CTGAACGGCCAAGTGA 2834
Db 7963 TGGCACTTAAGTTGGAGGCCGACAGATTGTTCGACGTCAAGAACGAGGACGGAGATGTCA 8022
Qy 2835 ATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACTCCACGTTGAAGGAAAAA 2894
Db 8023 TCGGGCACGCACTGGCCATGGAAGGAAAGGTAATGAAACCTCTGCACGTGAAAGGAACCA 8082
Qy 2895 TTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAGCATGTACGACTTGGAGT 2954
Db 8083 TCGACCACCCTGTGCTATCAAAGCTCAAATTTACCAAGTCGTCAGCATACGACATGGAGT 8142
Qy 2955 ACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTACACCAGCGACAAACCAC 3014
Db 8143 TCGCACAGTTGCCAGTCAACATGAGAAGTGAGGCATTCACCTACACCAGTGAACACCCCG 8202
Qy 3015 CGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAATGGGAGATTTACCGTAC 3074
Db 8203 AAGGATTCTATAACTGGCACCACGGAGCGGTGCAGTATAGTGGAGGTAGATTTACCATCC 8262
Qy 3075 CGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGACAACAGAGGCAGAG 3134
Db 8263 CTCGCGGAGTAGGAGGCAGAGGAGACAGCGGTCGTCCGATCATGGATAACTCCGGTCGGG 8322
Qy 3135 TTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGCGCTTTCAGTGGTCA 3194
Db 8323 TTGTCGCGATAGTCCTCGGTGGAGCTGATGAAGGAACACGAACTGCCCTTTCGGTCGTCA 8382
Qy 3195 CTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTCTGAACCGTGGT--- 3251
Db 8383 CCTGGAATAGTAAAGGGAAGACAATTAAGACGACCCCGGAAGGGACAGAAGAGTGGTCCG 8442
Qy 3252 ------CACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGACAAAC 3305
Db 8443 CAGCACCACTGGTCACGGCAATGTGTTTGCTCGGAAATGTGAGCTTCCCATGCGACCGCC 8502
Qy 3306 CACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAACGTCG 3365
Db 8503 CGCCCACATGCTATACCCGCGAACCTTCCAGAGCCCTCGACATCCTTGAAGAGAACGTGA 8562
Qy 3366 ACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAATGTCCATCAC---GCCGGC 3422
Db 8563 ACCATGAGGCCTACGATACCCTGCTCAATGCCATATTGCGGTGCGGATCGTCTGGCAGAA 8622
Qy 3423 CCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCCCGT 3482
Db 8623 GCAAAAGAAGCGTCACTGACGACTTTACCCTGACCAGCCCCTACTTGGGCACATGCTCGT 8682
Qy 3483 ATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACGAAT 3542
Db 8683 ACTGCCACCATACTGAACCGTGCTTCAGCCCTGTTAAGATCGAGCAGGTCTGGGACGAAG 8742
Qy 3543 CTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAGGCA 3602
Db 8743 CGGACGATAACACCATACGCATACAGACTTCCGCCCAGTTTGGATACGACCAAAGCGGAG 8802
Qy 3603 CTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAAGGAAG 3662
Db 8803 CAGCAAGCGCAAACAAGTACCGCTACATGTCGCTTGAGCAGGATCACACCGTTAAAGAAG 8862
Qy 3663 ACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCACAAAG 3722
Db 8863 GCACCATGGATGACATCAAGATTAGCACCTCAGGACCGTGTAGAAGGCTTAGCTACAAAG 8922
Qy 3723 GGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCACGAGCG 3782
Db 8923 GATACTTTCTCCTCGCAAAATGCCCTCCAGGGGACAGCGTAACGGTTAGCATAGTGAGTA 8982
Qy 3783 GAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTAGAG 3842
Db 8983 GCAACTCAGCAACGTCATGTACACTGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGG 9042
Qy 3843 AGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATCACT 3902
Db 9043 AAAAATATGATCTACCTCCCGTTCACGGTAAAAAAATTCCTTGCACAGTGTACGACCGTC 9102
Qy 3903 TGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAAGT 3962
Db 9103 TGAAAGAAACAACTGCAGGCTACATCACTATGCACAGGCCGGGACCGCACGCTTATACAT 9162
Qy 3963 CCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGTCA 4022
Db 9163 CCTACCTGGAAGAATCATCAGGGAAAGTTTACGCAAAGCCGCCATCTGGGAAGAACATTA 9222
Qy 4023 CCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGATGA 4082
Db 9223 CGTATGAGTGCAAGTGCGGCGACTACAAGACCGGAACCGTTTCGACCCGCACCGAAATCA 9282
Qy 4083 ACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGTCT 4142
Db 9283 CTGGTTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAACGAAGTGGGTCT 9342
Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9343 TCAACTCACCGGACTTGATCAGACATGACGACCACACGGCCCAAGGGAAATTGCATTTGC 9402
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9403 CTTTCAAGTTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAATGTAATAC 9462
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9463 ATGGCTTTAAACACATCAGCCTCCAATTAGATACAGACCACTTGACATTGCTCACCACCA 9522
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382

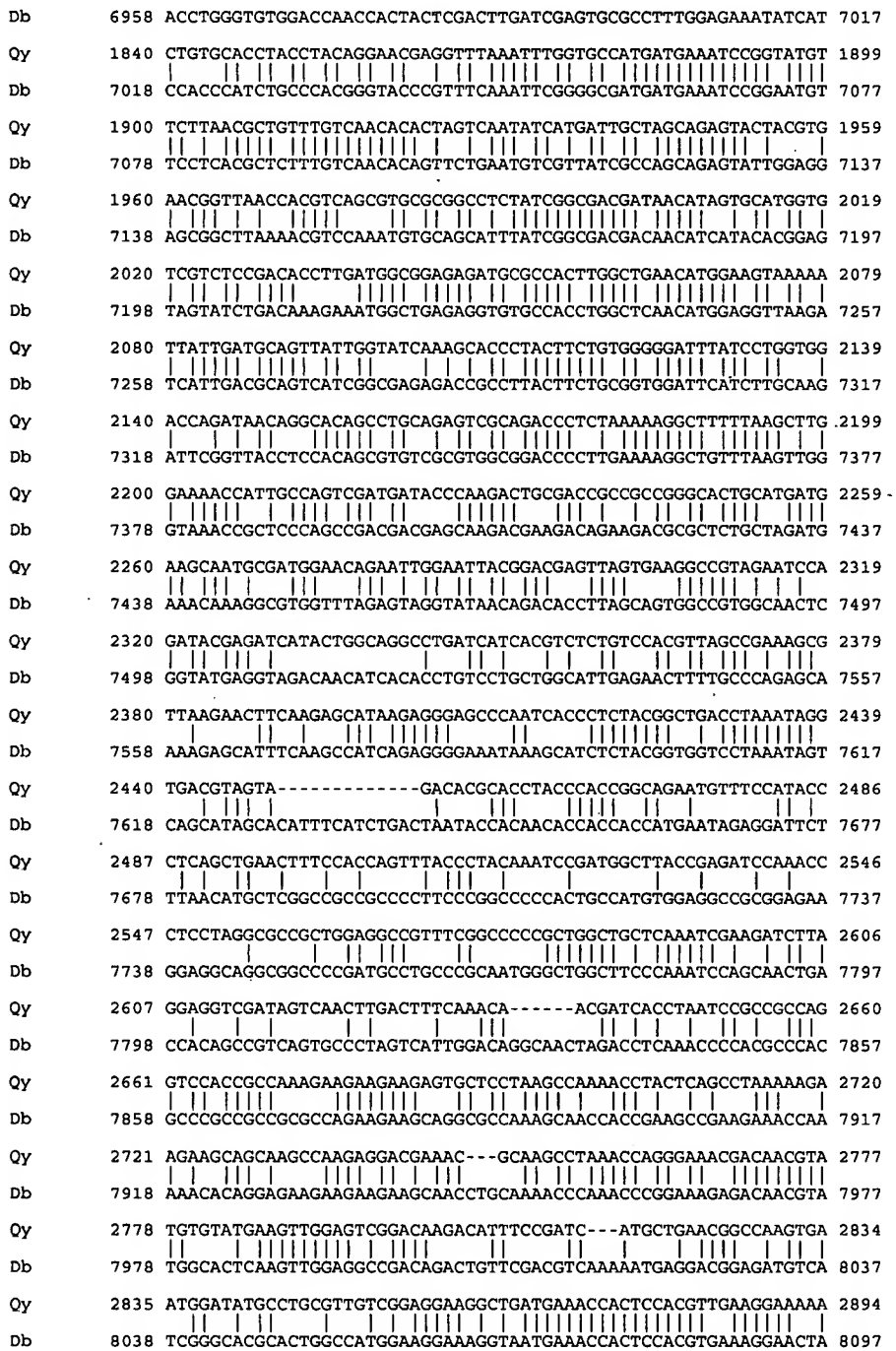
```
Db       9523 GGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCGTCGGAAAGACGGTCAGAA 9582
Qy       4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db       9583 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGAAATCATGAGCCAGTGAGGG 9642
Qy       4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db       9643 TCTATGCCCAAGAGTCAGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9702
Qy       4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db       9703 ATTACTACCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTACCGTGGCGA 9762
Qy       4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db       9763 TGATGATTGGCGTAACCGTTGCAGTGTTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9822
Qy       4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db       9823 CGCCATACGCCCTGGCCCCAAACGCCGTAATCCCAACTTCGCTGGCACTCTTGTGCTGCG 9882
Qy       4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db       9883 TTAGGTCGGCCAATGCTGAAACGTTCACCGAGACCATGAGTTACTTGTGGTCGAACAGTC 9942
Qy       4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
Db       9943 AGCCGTTCTTCTGGGTCCAGTTGTGCATACCTTTGGCCGCTTTCATCGTTCTAATGCGCT 10002
Qy       4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
Db      10003 GCTGCTCCTGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10062
Qy       4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
Db      10063 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10122
Qy       4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db      10123 AAAGGGCAGGGTATGCCCCGCTCAATTTGGAGATCACTGTCATGTCCTCGGAGGTTTTGC 10182
Qy       4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db      10183 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCAAAAA 10242
Qy       5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db      10243 TCAAATGCTGCGGCTCCTTGGAATGTCAGCCGGCCGCTCATGCAGACTATACCTGCAAGG 10302
Qy       5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db      10303 TCTTCGGAGGGGTCTACCCCTTTATGTGGGGAGGAGCGCAATGTTTTTGCGACAGTGAGA 10362
Qy       5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db      10363 ACAGCCAGATGAGTGAGGCGTACGTCGAACTGTCAGCAGATTGCGCGTCTGACCACGCGC 10422
Qy       5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db      10423 AGGCGATTAAGGTGCACACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10482
Qy       5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db      10483 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACTTGA 10542
Qy       5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db      10543 AAGTCATAGCTGGACCAATTTCAGCATCGTTTACGCCATTCGATCATAAGGTCGTTATCC 10602
Qy       5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db      10603 ATCGCGGCCTGGTGTACAACTATGACTTCCCGGAATATGGAGCGATGAAACCAGGAGCGT 10662
```

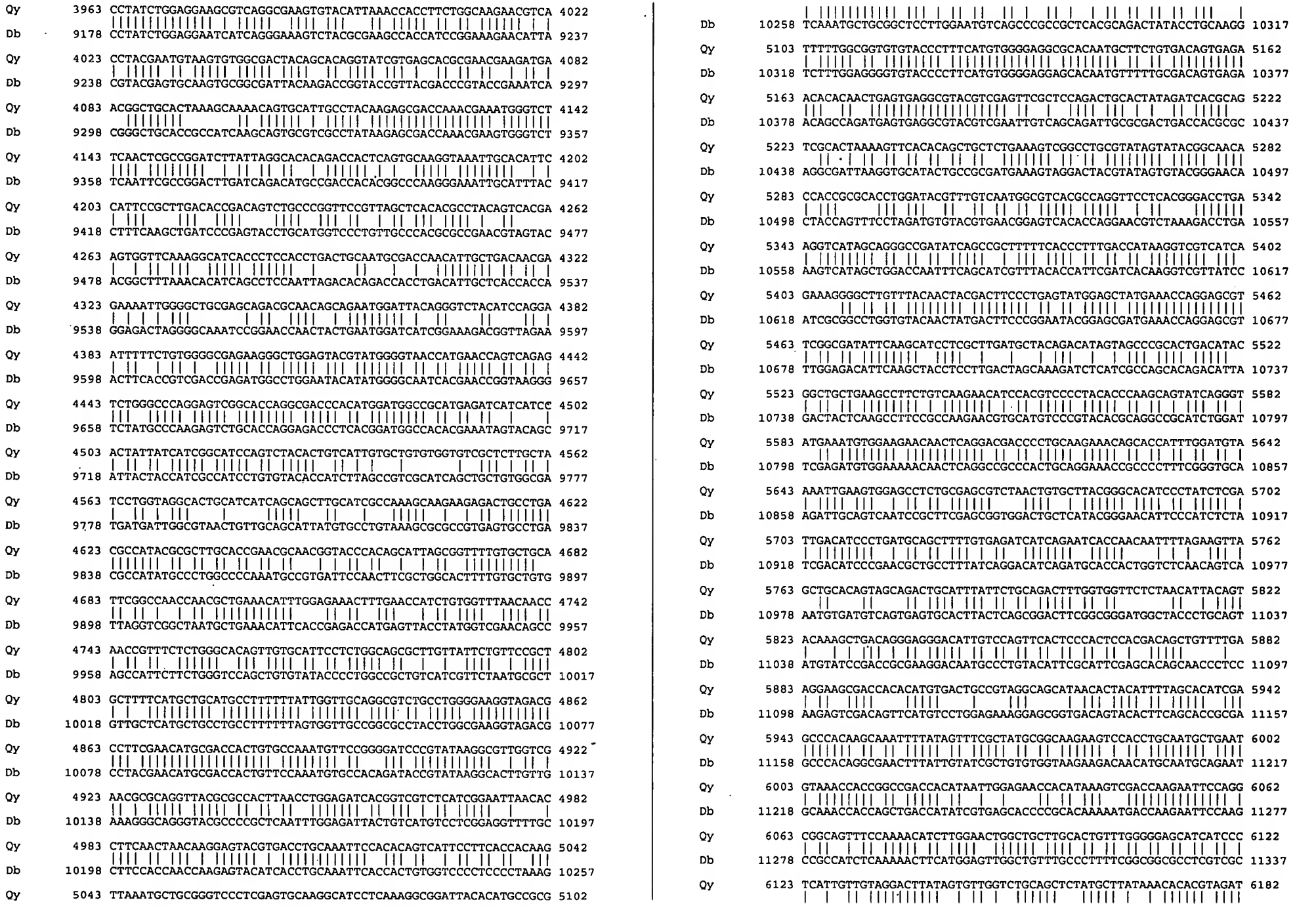
```
Qy       5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db      10663 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAGGATCTCATCGCCAGCACAGACATTA 10722
Qy       5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db      10723 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCAGGAT 10782
Qy       5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db      10783 TTGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCACCTTTCGGGTGTA 10842
Qy       5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db      10843 AGATTGCAGTAAATCCGCTCCGAGCGGTGGACTGTTCATACGGGAACATTCCCATTTCTA 10902
Qy       5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db      10903 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10962
Qy       5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db      10963 AATGTGAAGTCAGTGAGTGCACTTATTCAGCAGACTTCGGCGGGATGGCCACCCTGCAGT 11022
Qy       5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db      11023 ATGTATCCGACCGCGAAGGTCAATGCCCCGTACATTCGCATTCGAGCACAGCAACTCTCC 11082
Qy       5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db      11083 AAGAGTCGACAGTACATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTTAGCACCGCGA 11142
Qy       5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db      11143 GTCCACAGGCGAACTTTATCGTATCGCTGTGTGGGAAGAAGACAACATGCAATGCAGAAT 11202
Qy       6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db      11203 GTAAACCACCAGCTGACCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTTCAAG 11262
Qy       6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db      11263 CCGCCATCTCAAAAACATCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11322
Qy       6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
Db      11323 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11382
Qy       6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
Db      11383 GACCGCTACGCCCCAATGATCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11442
Qy       6242 GCATAATGCCACGCGCCGCT 6261
Db      11443 GCATAATGCATCAGGCTGGT 11462
```

```
RESULT 5
AAV33474
ID   AAV33474 standard; cDNA; 11717 BP.
XX
AC   AAV33474;
XX
DT   10-DEC-1998  (first entry)
XX
DE   Girdwood S.A.virus cDNA sequence.
XX
KW   Bone marrow cell; alphavirus; Girdwood S.A.virus; cytokine; capsid;
KW   reverse transcriptase-PCR; nsP1; nsP2; nsP3; nsP4; E3; E2; 6K; E1;
KW   growth hormone; growth factor; interleukin; chemokine; enzyme;
KW   ribozyme; antisense oligonucleotide; ss.
XX
```

```
OS   Girdwood S.A. virus.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..59
FT                   /*tag=  a
FT   CDS             60..7616
FT                   /*tag=  b
FT                   /product= "Non-structural polyprotein"
FT                   /note= "Internal stop codon present at nucleotides
FT                        5763-5765"
FT   CDS             7662..11399
FT                   /*tag=  c
FT                   /product= "Structural polyprotein"
FT   3'UTR           11400..11717
FT                   /*tag=  d
XX
PN   WO9836779-A2.
XX
PD   27-AUG-1998.
XX
PF   18-FEB-1998;   98WO-US02945.
XX
PR   19-FEB-1997;   97US-0801263.
XX
PA   (UYNC-) UNIV NORTH CAROLINA.
XX
PI   Davis NL,  Johnston RE,  Simpson DA;
XX
DR   WPI; 1998-495361/42.
DR   P-PSDB; AAW70462, AAW70463, AAW70473, AAW70474, AAW70475, AAW70476,
DR   AAW70477, AAW70478, AAW70479.
XX
PT   Expression of heterologous RNA in bone marrow cells - using a
PT   recombinant alphavirus comprising a promoter operable in bone marrow
PT   cells operably associated with heterologous RNA.
XX
PS   Example 2; Fig 3A-3C; 68pp; English.
XX
CC   The invention provides a method for introducing and expressing
CC   heterologous RNA in bone marrow cells using alphavirus vectors.  The
CC   Girdwood S.A.virus is an alphavirus whose cDNA sequence was determined
CC   from uncloned reverse transcriptase-PCR reaction fragments amplified
CC   from the virion RNA.  The Girdwood S.A.virus cDNA encodes nonstructural
CC   and structural polyproteins.  The nonstructural polyprotein is,
CC   presumably, post-translationally modified into four different protein
CC   products, namely nsP1 (AAW70462), nsP2 (AAW70463), nsP3 (AAW70473), nsP4
CC   (AAW70474) proteins.  The structural polyprotein is, presumably, post-
CC   translationally modified into five different protein products, namely
CC   capsid (AAW70475), E3 (AAW70476), E2 (AAW70477), 6K (AAW70478), E1
CC   (AAW70479) proteins.  The Girdwood S.A.virus cDNA was used in the method
CC   of the invention.  The inventors claim the transformed bone marrow cells
CC   are useful for expressing a protein or peptide suitable for protecting
CC   the subject against a disease such as a microbial, bacterial, protozoal,
CC   parasitic or viral disease.  The transformed bone marrow cells are also
CC   claimed to be useful for expressing proteins and peptides such as
CC   hormones, growth hormones, growth factors, interleukins, cytokines,
CC   chemokines, enzymes, ribozymes or antisense oligonucleotides.
XX
SQ   Sequence 11717 BP; 3297 A; 3130 C; 2927 G; 2360 T; 3 other;

  Query Match           36.8%;  Score 2386.4;  DB 19;  Length 11717;
  Best Local Similarity  64.5%;  Pred. No. 0;
  Matches 3745;  Conservative    0;  Mismatches 1991;  Indels   74;  Gaps    9;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
              ||||||     |  |  || | ||| || || ||  |||| |||||||  || ||| | |
Db       5761 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5820

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
              ||  | | || || |  ||   | | ||   |||||| |  ||||| ||    || |  |
Db       5821 AAATGGAGTCCGTTCTGCAGAATCAGCTTACAGAACCGACCTTGGAGCGCAATGTTCTGG 5880

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
```

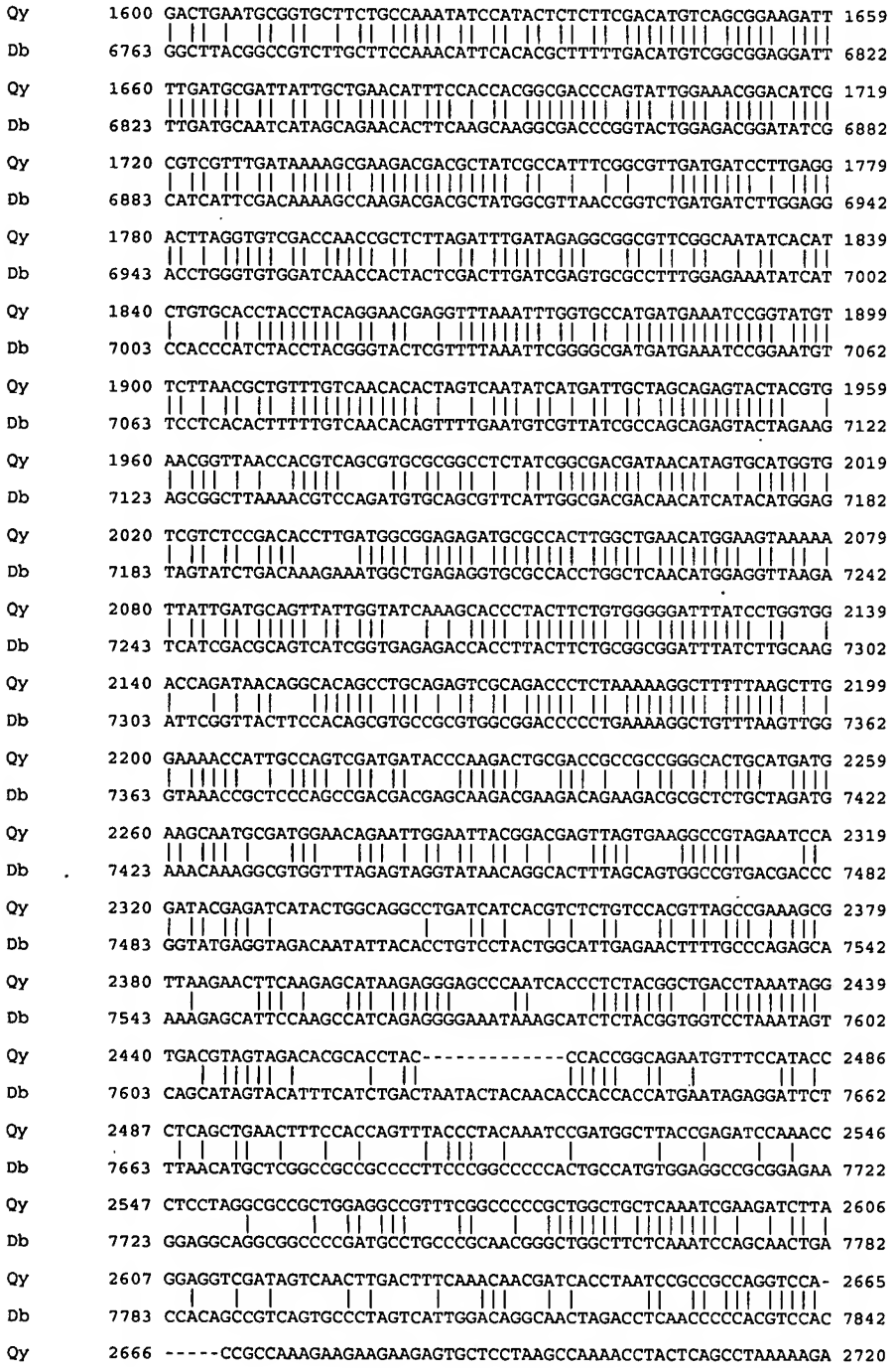
```
              | |    |||||||||   |||||       | |||| || | |  |    |    |
Db       5881 AAAGAATCTACGCCCCGGTGCTCGACACGTCGAAAGAGGAACAGCTCAAACTCAGGTACC 5940

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
              |  |      || | |||| || | |||||||||| || ||  |||||||||||||  ||
Db       5941 AGATGATGCCCACCGAAGCCAACAAAAGCAGGTACCAGTCTAGAAAAGTAGAAAATCAGA 6000

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
              |||| || ||  | |||||||  |||| ||  |  |    |||   ||  |   ||
Db       6001 AAGCCATAACCACTGAGCGACTGCTTTCAGGGCTACGACTGTATAACTCTGCCACAG--- 6057

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
              |||     || || || |   |||  || ||   ||||   || ||     |  |||
Db       6058 ATCAGCCAGAATGCTATAAGATCACCTACCCGAAACCATCGTATTCCAGCAGTGTACCGG 6117

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
              |  |   |     || || | || ||      ||||||    | | || |||||||
Db       6118 CGAACTACTCTGACCCAAAGTTTGCTGTAGCTGTTTGCAACAACTATCTGCATGAGAATT 6177

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
              |||| || |||||   ||||  || || || || |||||||| ||  | || ||||| |
Db       6178 ACCCGACGGTAGCATCTTATCAGATCACCGACGAGTACGATGCTTACTTGGATATGGTAG 6237

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
              ||||  ||      || |||||||| || |||||||| || || || || ||||| ||||
Db       6238 ACGGGACAGTCGCTTGCCTAGATACTGCAACTTTTTGCCCCGCCAAGCTTAGAAGTTACC 6297

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
              | || |  ||   |||  . ||| | |   |  || || |||||  | || ||||
Db       6298 CGAAAAGACACGAGTATAGAGCCCCAAACACTCGCAGTGCGGTTCCATCAGCGATGCAGA 6357

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
              | || || ||||| ||  |   ||| || |||||||| || |||||||| || |||||||
Db       6358 ACACGTTGCAAAACGTGCTCATTGCCGCGACTAAAAGAAACTGCAACGTCACACAAATGC 6417

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
              | ||||| ||     | || || ||| |||| || ||||| || ||    ||||| ||||
Db       6418 GTGAATTGCCAACACTGGACTCAGCGACATTCAACGTTGAATGCTTTCGAAAATATGCAT 6477

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
              | ||||| ||||| |||||   |||  |   || || ||| || | |||| |||  ||
Db       6478 GTAATGACGAGTATTGGGAGGAGTTTGCCCGAAAGCCAATTAGGATCACTACTGAGTTCG 6537

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
              ||||   ||| ||| | |  |||||||| || || || || ||| |||| || || || |
Db       6538 TTACCGCATACGTGGCCAGACTGAAAGGCCCTAAGGCCGCCGCACTGTTCGCAAAGACGC 6597

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
              |||||  |    || ||||| ||  | || ||||||   |||||||||||  | || ||||
Db       6598 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGGTTCGTCATGGACATGAAAAGAG 6657

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
              | || |||||||| || |||||||||| ||||| ||  | || || || || || | |
Db       6658 ACGTGAAAGTTACACCTGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGCTAC 6717

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
              | || ||||| ||||| || |||||||||| |||||||||| ||||| ||||| || |
Db       6718 AAGCCGCAGAACCCCTGGCGACCGCTTACCTGTGCGGGATCCACCGGGAGTTAGTGCGCA 6777

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
              | || |  || ||  | || || || || || || || || || |||||||| ||||| || |
Db       6778 GGCTTACAGCCGTCTTGCTACCCAACATTCACACGCTTTTTGACATGTCGGCGGAGGACT 6837

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
              ||||||| || || || ||||| ||| | || |||||  ||| |||| ||||| ||||
Db       6838 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGTGACCCGGTACTGGAGACGGATATCG 6897

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
              | ||||| || |||||| |||||||||||||| ||  |  | |  |||||||| | || |
Db       6898 CCTCGTTCGACAAAAGCCAAGACGACGCTATGGCGTTAACTGGCCTGATGATCTTGGAAG 6957

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
              || | ||||| |||||||| ||  | || ||||| |||   || || ||  | ||  |||
```

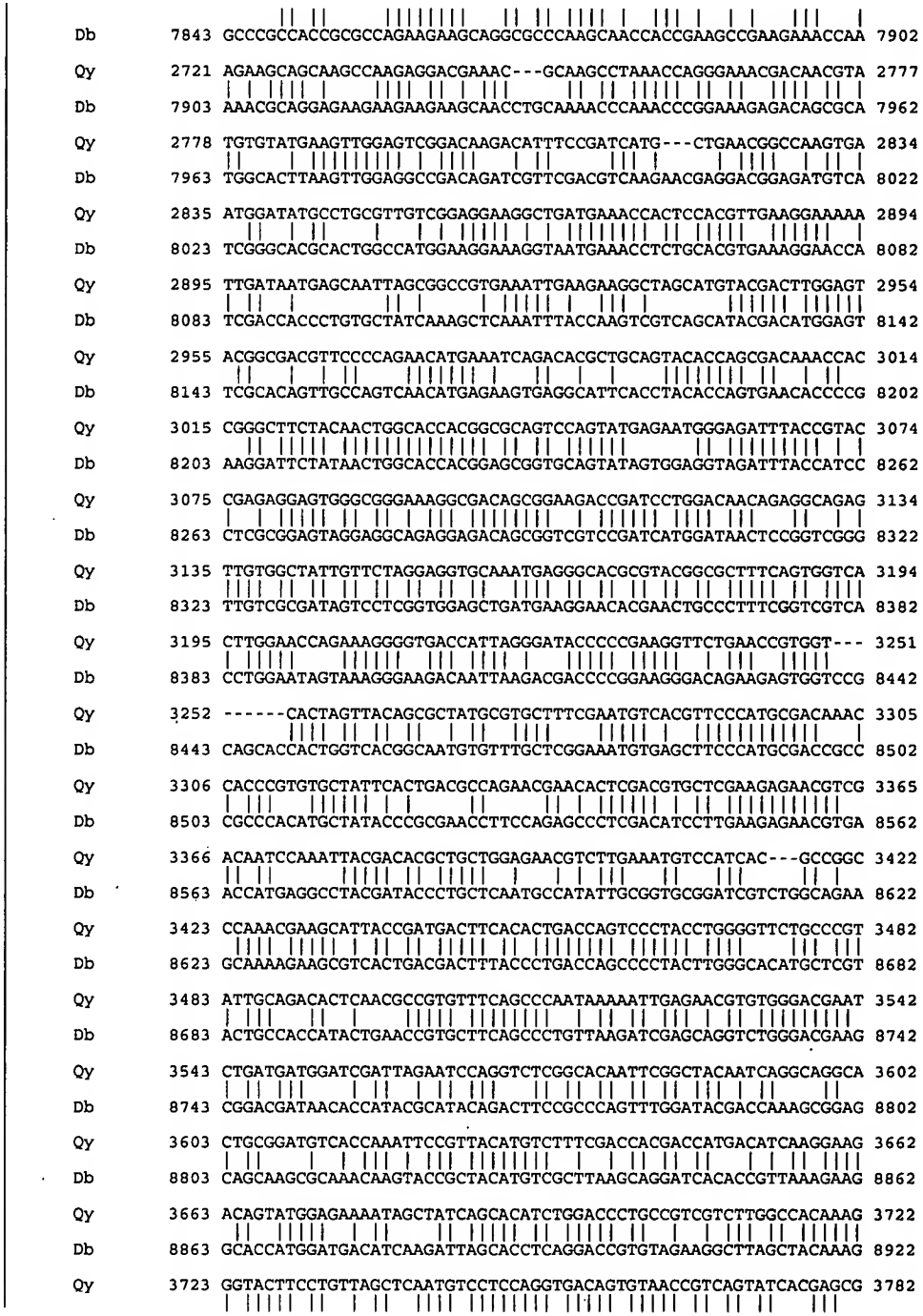

Db 6958 ACCTGGGTGTGGACCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 7017
Qy 1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
Db 7018 CCACCCATCTGCCCACGGGTACCCGTTTCAAATTCGGGGCGATGATGAAATCCGGAATGT 7077
Qy 1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
Db 7078 TCCTCACGCTCTTTGTCAACACAGTTCTGAATGTCGTTATCGCCAGCAGAGTATTGGAGG 7137
Qy 1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
Db 7138 AGCGGCTTAAAACGTCCAAATGTGCAGCATTTATCGGCGACGACAACATCATACACGGAG 7197
Qy 2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
Db 7198 TAGTATCTGACAAAGAAATGGCTGAGAGGTGTGCCACCTGGCTCAACATGGAGGTTAAGA 7257
Qy 2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139
Db 7258 TCATTGACGCAGTCATCGGCGAGAGACCGCCTTACTTCTGCGGTGGATTCATCTTGCAAG 7317
Qy 2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199
Db 7318 ATTCGGTTACCTCCACAGCGTGTCGCGTGGCGGACCCCTTGAAAAGGCTGTTTAAGTTGG 7377
Qy 2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
Db 7378 GTAAACCGCTCCCAGCCGACGACGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 7437
Qy 2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319
Db 7438 AAACAAAGGCGTGGTTTAGAGTAGGTATAACAGACACCTTAGCAGTGGCCGTGGCAACTC 7497
Qy 2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379
Db 7498 GGTATGAGGTAGACAACATCACACCTGTCCTGCTGGCATTGAGAACTTTTGCCCAGAGCA 7557
Qy 2380 TTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGG 2439
Db 7558 AAAGAGCATTTCAAGCCATCAGAGGGGAAATAAAGCATCTCTACGGTGGTCCTAAATAGT 7617
Qy 2440 TGACGTAGTA----------GACACGCACCTACCCACCGGCAGAATGTTTCCATACC 2486
Db 7618 CAGCATAGCACATTTCATCTGACTAATACCACAACACCACCACCATGAATAGAGGATTCT 7677
Qy 2487 CTCAGCTGAACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACC 2546
Db 7678 TTAACATGCTCGGCCGCCGCCCCTTCCCGGCCCCCACTGCCATGTGGAGGCCGCGGAGAA 7737
Qy 2547 CTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTA 2606
Db 7738 GGAGGCAGGCGGCCCCGATGCCTGCCCGCAATGGGCTGGCTTCCCAAATCCAGCAACTGA 7797
Qy 2607 GGAGGTCGATAGTCAACTTGACTTTCAAACA------ACGATCACCTAATCCGCCGCCAG 2660
Db 7798 CCACAGCCGTCAGTGCCCTAGTCATTGGACAGGCAACTAGACCTCAAACCCCACGCCCAC 7857
Qy 2661 GTCCACCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGA 2720
Db 7858 GCCCGCCGCCGCGCCAGAAGAAGCAGGCGCCAAAGCAACCACCGAAGCCGAAGAAACCAA 7917
Qy 2721 AGAAGCAGCAAGCCAAGAGGACGAAAC---GCAAGCCTAAACCAGGGAAACGACAACGTA 2777
Db 7918 AAACACAGGAGAAGAAGAAGAAGCAACCTGCAAAACCCAAACCCGGAAAGAGACAACGTA 7977
Qy 2778 TGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATC---ATGCTGAACGGCCAAGTGA 2834
Db 7978 TGGCACTCAAGTTGGAGGCCGACAGACTGTTCGACGTCAAAAATGAGGACGGAGATGTCA 8037
Qy 2835 ATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACTCCACGTTGAAGGAAAAA 2894
Db 8038 TCGGGCACGCACTGGCCATGGAAGGAAAGGTAATGAAACCACTCCACGTGAAAGGAACTA 8097
Qy 2895 TTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAGCATGTACGACTTGGAGT 2954
Db 8098 TTGACCACCCTGTGCTATCAAAGCTCAAATTCACCAAGTCGTCAGCATACGACATGGAGT 8157
Qy 2955 ACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTACACCAGCGACAAACCAC 3014
Db 8158 TCGCACAGTTGCCGGTCAACATGAGAAGTGAGGCGTTCACCTACACCAGCGAACACCCTG 8217
Qy 3015 CGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAATGGGAGATTTACCGTAC 3074
Db 8218 AAGGGTTTTACAACTGGCACCACGGAGCGGTGCAGTATAGTGGAGGTAGATTTACCATCC 8277
Qy 3075 CGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGACAACAGAGGCAGAG 3134
Db 8278 CCCGCGGAGTAGGAGGCAGAGGAGACAGTGGTCGTCCGATTATGGATAACTCAGGCCGGG 8337
Qy 3135 TTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGCGCTTTCAGTGGTCA 3194
Db 8338 TTGTCGCGATAGTCCTCGGAGGGGCTGATGAGGGAACAAGAACTGCCCTTTCGGTCGTCA 8397
Qy 3195 CTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTCTGAACCGTGGT--- 3251
Db 8398 CCTGGAATAGCAAAGGGAAGACAATCAAGACAACCCCGGAAGGGACAGAAGAGTGGTCTG 8457
Qy 3252 ------CACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGACAAAC 3305
Db 8458 CAGCACCACTGGTCACGGCCATGTGCTTGCTTGGAAACGTGAGCTTCCCATGCAATCGCC 8517
Qy 3306 CACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAACGTCG 3365
Db 8518 CGCCCACATGCTACACCCGCGAACCATCCAGAGCTCTTGACATCCTTGAAGAGAACGTGA 8577
Qy 3366 ACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAAT---GTCCATCACGCCGGC 3422
Db 8578 ACCACGAGGCCTACGACACCCTGCTCAACGCCATATTGCGGTGCGGATCGTCCGGCAGAA 8637
Qy 3423 CCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCCCGT 3482
Db 8638 GCAAAAGAAGCGTCACTGACGACTTTACCTTGACCAGCCCGTACTTGGGCACATGCTCGT 8697
Qy 3483 ATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACGAAT 3542
Db 8698 ACTGTCACCATACTGAACCGTGCTTTAGCCCGATTAAGATCGAGCAGGTCTGGGATGAAG 8757
Qy 3543 CTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAGGCA 3602
Db 8758 CGGACGACAACACCATACGCATACAGACTTCCGCCCAGTTTGGATACGACCAAAGCGGAG 8817
Qy 3603 CTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAAGGAAG 3662
Db 8818 CAGCAAGCTCAAATAAGTACCGCTACATGTCGCTCGAGCAGGATCATACCGTCAAAGAAG 8877
Qy 3663 ACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCACAAAG 3722
Db 8878 GCACTATGGATGACATCAAGATCAGCACCTCAGGACCGTGTAGAAGGCTTAGCTACAAAG 8937
Qy 3723 GGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCACGAGCG 3782
Db 8938 GATACTTTCTCCTCGCGAAGTGTCCTCCAGGGGACAGCGTAACGGTTAGTATAGCGAGTA 8997
Qy 3783 GAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTAGAG 3842
Db 8998 GCAACTCAGCAACGTCATGCACAATGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGG 9057
Qy 3843 AGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATCACT 3902
Db 9058 AAAAATATGACCTACCTCCCGTTCACGGTAAGAAGATTCCTTGCACAGTGTACGACCGTC 9117
Qy 3903 TGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAAGT 3962
Db 9118 TGAAAGAAACAACCGCCGGCTACATCACTATGCACAGGCCGGGACCGCACGCCTATACGT 9177

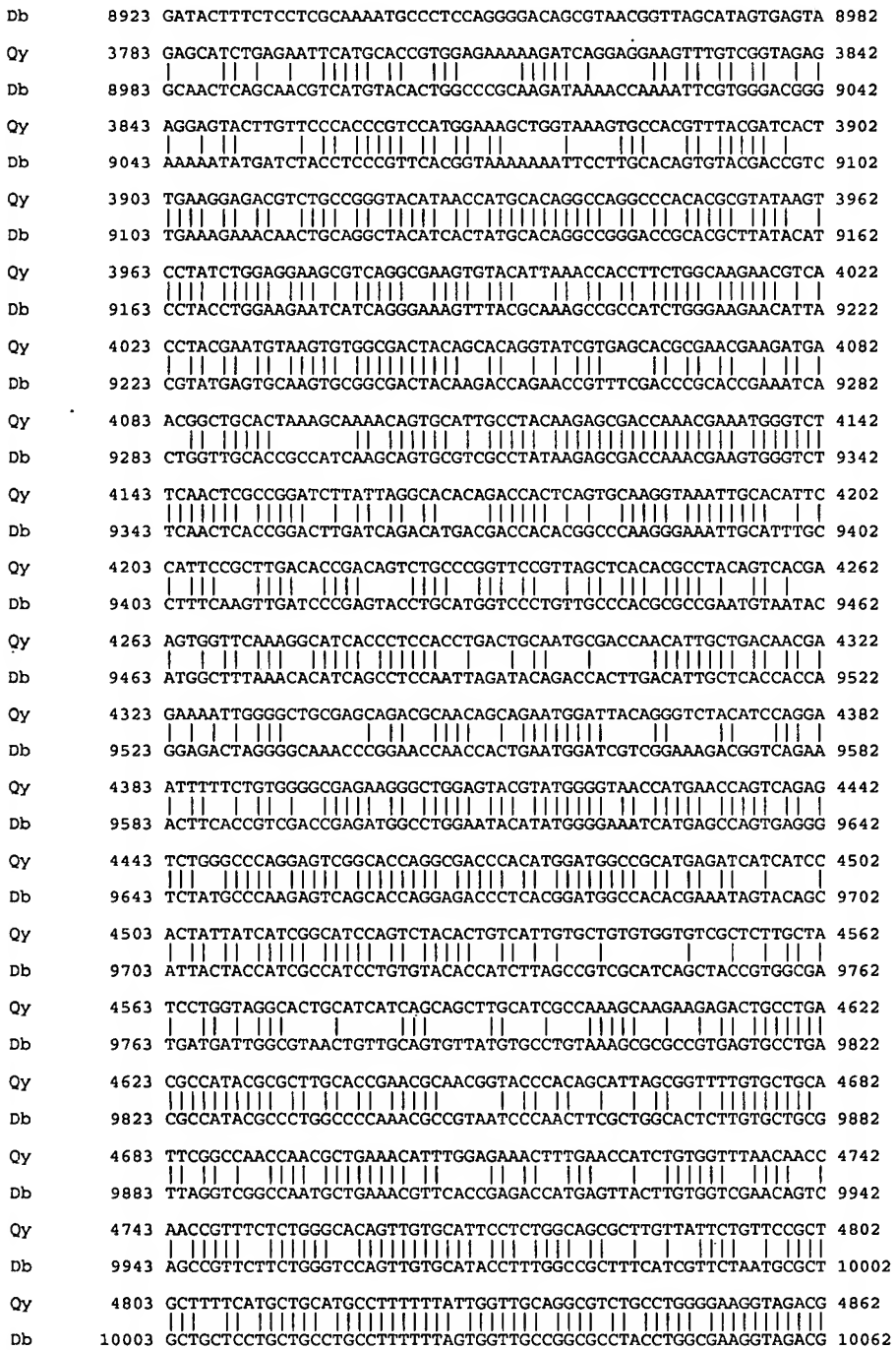


```
Db      11338 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11397

Qy       6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
              ||| |         || ||| |  | | | ||||||||||||||||||||||||  |  ||
Db      11398 GACCGCTACGCCCCAATGACCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11457

Qy       6242 GCATAATGCCACGCGCCGCT--------------------------------TGACACT 6268
              |||||||||  |  || | |                                ||||
Db      11458 GCATAATGCATCAGGCTGGTATATTAGATCCCCGCTTACCGCGGGCAATATAGCAACACC 11517

Qy       6269 AAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGT 6318
              ||||||||| |||||||||||||||| |||||||||||||| |||||||
Db      11518 AAAACTCGACGTATTTCCGAGGAAGCGCAGTGCATAATGCTGCGCAGTGT 11567


RESULT 6
AAC64506
ID   AAC64506 standard; DNA; 11703 BP.
XX
AC   AAC64506;
XX
DT   14-FEB-2001  (first entry)
XX
DE   Alphavirus SinDCChiron virus nucleotide sequence.
XX
KW   Alphavirus; SinDCChiron virus; SinChironLP virus; immune response;
KW   infection; human dendritic cell; immunostimulatory; cytostatic;
KW   virucide; fungicide; antibacterial; antiparasitic; vaccine; cancer;
KW   pathogen; antigen presenting cell; ds.
XX
OS   Alphavirus.
XX
PN   WO200061772-A2.
XX
PD   19-OCT-2000.
XX
PF   14-APR-2000; 2000WO-US10722.
XX
PR   14-APR-1999;   99US-0129498.
PR   09-AUG-1999;   99US-0148086.
PR   22-MAR-2000; 2000US-0191363.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Polo JM,  Dubensky TW,  Frolov I,  Gardner JP,  Otten G,  Barnett S;
PI   Driver DA;
XX
DR   WPI; 2000-619231/59.
XX
PT   New alphavirus that infects human dendritic cells for use in generating
PT   an immune response to pathogenic agents such as bacteria, viruses,
PT   fungi, parasites and cancer and for biological assays -
XX
PS   Claim 9; Fig 2B; 83pp; English.
XX
CC   The present invention describes an isolated alphavirus (AV) which
CC   infects human dendritic cells and is not of American Type Culture
CC   Collection (ATCC) number VR-2526. AAC64506 and AAC64507 represent the
CC   nucleotide sequence of the specifically claimed SinDCChiron virus and
CC   SinChironLP virus. The new AVs have immunostimulatory, cytostatic,
CC   virucide, fungicide, antibacterial and antiparasitic activities and
CC   can be used in vaccines. The AVs are used to infect dendritic cells,
CC   preferably human cells. A heterologous sequence can be introduced and
CC   expressed in human macrophages or antigen presenting cells in vivo and
CC   in vitro, for use in biological assays. The AV-based vector systems
CC   are used to generate an immune response to cancer or a pathogenic
CC   agent, such as, bacteria, fungi, parasites or viruses. The AV can be
CC   used to infect human dendritic cells, macrophages or antigen
CC   presenting cells that previously could not be infected using an AV or
CC   AV variant. The AV vectors are targeted directly to antigen presenting
CC   cells.
XX
SQ   Sequence 11703 BP; 3310 A; 3049 C; 2910 G; 2434 T; 0 other;
```

```
  Query Match             36.8%;  Score 2383.6;  DB 21;  Length 11703;
  Best Local Similarity   64.6%;  Pred. No. 0;
  Matches 3695;  Conservative    0;  Mismatches 1984;  Indels   41;  Gaps    8;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
              ||||||     |  |  || | ||| || || ||  |||| |||||||  || ||| | |
Db       5746 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5805

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
              || |||| || ||  |  ||      | ||   ||||||  |  |||||  ||    ||||  |
Db       5806 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTCCTGG 5865

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
              | |    | | ||||||   |||||        | ||||  ||    |  |    |     |
Db       5866 AAAGAATTCATGCCCCGGTGCTCGACACGTCGAAAGAGGAACAACTCAAACTCAGGTACC 5925

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
              |  |      || | ||||  || | ||| |||||  || || || ||||||||||||  ||
Db       5926 AGATGATGCCCACCGAAGCCAACAAAAGTAGGTACCAGTCTCGTAAAGTAGAAAATCAGA 5985

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
              |||| || ||  | ||||||||  | || ||| |  |    |||  ||  |   ||
Db       5986 AAGCCATAACCACTGAGCGACTACTGTCAGGACTACGACTGTATAACTCTGCCACAG--- 6042

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
              |||     || || || |   |||  |||||   |||| | |||||     |  |||
Db       6043 ATCAGCCAGAATGCTATAAGATCACCTATCCGAAACCATTGTACTCCAGTAGCGTACCGG 6102

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
                |  |   |  | || || |||| ||       || |||    |  | || |||||  |
Db       6103 CGAACTACTCCGATCCACAGTTCGCTGTAGCTGTCTGTAACAACTATCTGCATGAGAACT 6162

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
              | || ||||||||   ||||  || || || || |||||||| ||  |  | |||||  |
Db       6163 ATCCGACAGTAGCATCTTATCAGATTACTGACGAGTACGATGCTTACTTGGATATGGTAG 6222

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
              ||||  ||      ||| || ||||| || || || || || ||||| || ||||| ||||
Db       6223 ACGGGACAGTCGCCTGCCTGGATACTGCAACCTTCTGCCCCGCTAAGCTTAGAAGTTACC 6282

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
              | || || |||   |||     .||  | ||  |    || || |||||  | || ||||
Db       6283 CGAAAAAACATGAGTATAGAGCCCCGAATATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6342

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
              | ||  ||||||||||  |  ||| || |||||||| |||||||||| || || ||||
Db       6343 ACACGCTACAAAATGTGCTCATTGCCGCAACTAAAAGAAATTGCAACGTCACGCAGATGC 6402

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
              | ||| | ||      | || || ||| |||| ||||| || || ||     ||||| ||||
Db       6403 GTGAACTGCCAACACTGGACTCAGCGACATTCAATGTCGAATGCTTTCGAAAATATGCAT 6462

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
              | ||||| ||||| |||||     ||        || || ||| || | || || |||  |
Db       6463 GTAATGACGAGTATTGGGAGGAGTTCGCTCGGAAGCCAATTAGGATTACCACTGAGTTTG 6522

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
              | ||   ||||||  | |  |||||||| || || || || ||| | ||||| || ||
Db       6523 TCACCGCATATGTAGCTAGACTGAAAGGCCCTAAGGCCGCCGCACTATTTGCAAAGACGT 6582

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
              |||||  |    || ||||| ||  | || ||||||  |||||||||||  | || || |
Db       6583 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGGG 6642

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
              | || |||||||| || ||||||||||| ||||| ||  |  || || || || || || | |
Db       6643 ACGTGAAAGTTACACCAGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6702

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
              | || ||||| |||||| || || |||||| | |||||||| || ||||||||||| ||||
Db       6703 AAGCCGCAGAACCCCTGGCGACTGCTTACTTATGCGGGATTCACCGGGAATTAGTGCGTA 6762
```





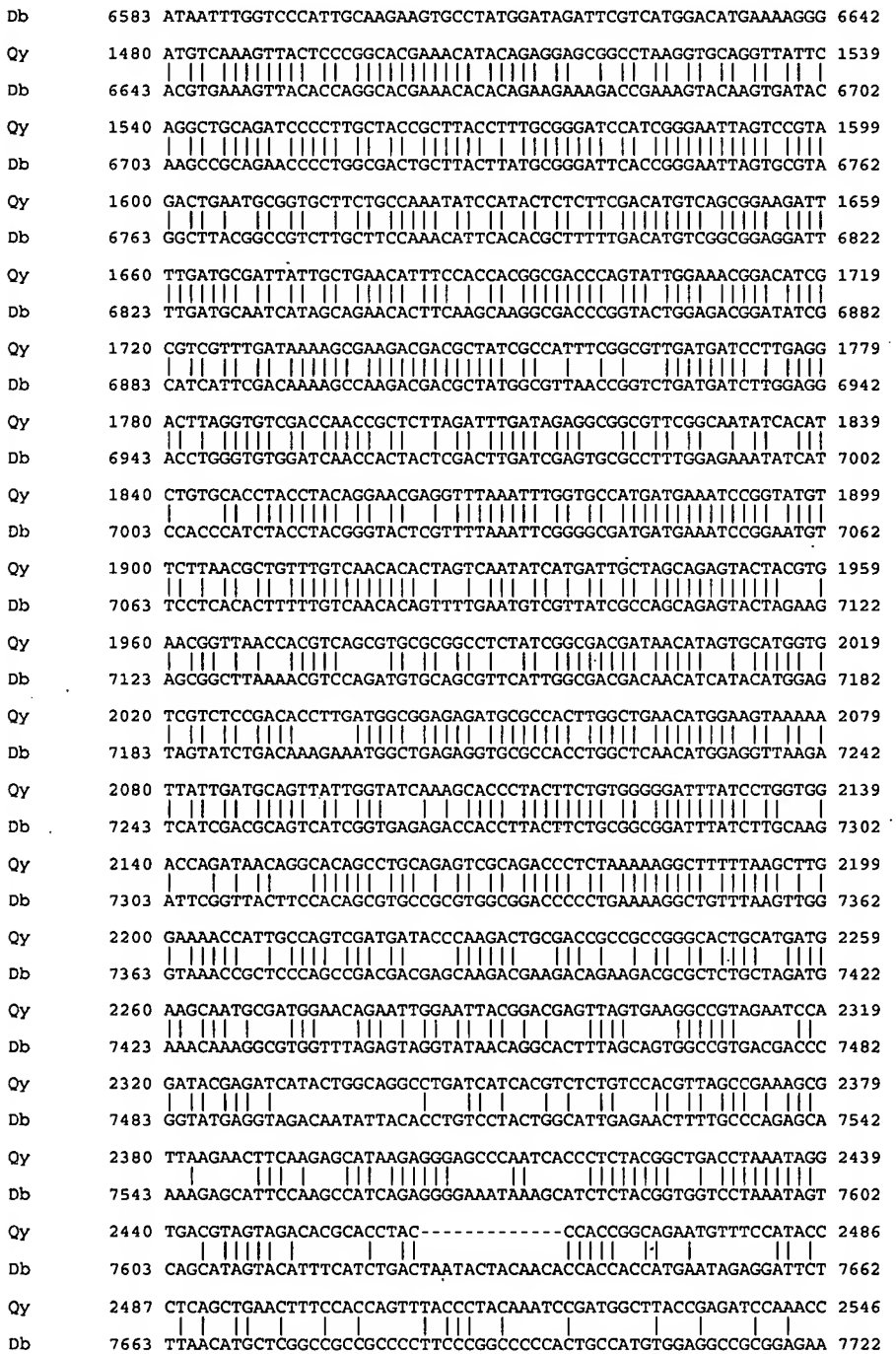

Db 8923 GATACTTTCTCCTCGCAAAATGCCCTCCAGGGGACAGCGTAACGGTTAGCATAGTGAGTA 8982
Qy 3783 GAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTAGAG 3842
Db 8983 GCAACTCAGCAACGTCATGTACACTGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGG 9042
Qy 3843 AGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATCACT 3902
Db 9043 AAAAATATGATCTACCTCCCGTTCACGGTAAAAAAATTCCTTGCACAGTGTACGACCGTC 9102
Qy 3903 TGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAAGT 3962
Db 9103 TGAAAGAAACAACTGCAGGCTACATCACTATGCACAGGCCGGGACCGCACGCTTATACAT 9162
Qy 3963 CCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGTCA 4022
Db 9163 CCTACCTGGAAGAATCATCAGGGAAAGTTTACGCAAAGCCGCCATCTGGGAAGAACATTA 9222
Qy 4023 CCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGATGA 4082
Db 9223 CGTATGAGTGCAAGTGCGGCGACTACAAGACCAGAACCGTTTCGACCCGCACCGAAATCA 9282
Qy 4083 ACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGTCT 4142
Db 9283 CTGGTTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAACGAAGTGGGTCT 9342
Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9343 TCAACTCACCGGACTTGATCAGACATGACGACCACACGGCCCAAGGGAAATTGCATTTGC 9402
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9403 CTTTCAAGTTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAATGTAATAC 9462
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9463 ATGGCTTTAAACACATCAGCCTCCAATTAGATACAGACCACTTGACATTGCTCACCACCA 9522
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9523 GGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCGTCGGAAAGACGGTCAGAA 9582
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9583 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGAAATCATGAGCCAGTGAGGG 9642
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db 9643 TCTATGCCCAAGAGTCAGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9702
Qy 4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db 9703 ATTACTACCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTACCGTGGCGA 9762
Qy 4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db 9763 TGATGATTGGCGTAACTGTTGCAGTGTTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9822
Qy 4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db 9823 CGCCATACGCCCTGGCCCCAAACGCCGTAATCCCAACTTCGCTGGCACTCTTGTGCTGCG 9882
Qy 4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db 9883 TTAGGTCGGCCAATGCTGAAACGTTCACCGAGACCATGAGTTACTTGTGGTCGAACAGTC 9942
Qy 4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
Db 9943 AGCCGTTCTTCTGGGTCCAGTTGTGCATACCTTTGGCCGCTTTCATCGTTCTAATGCGCT 10002
Qy 4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
Db 10003 GCTGCTCCTGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10062
Qy 4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
Db 10063 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10122
Qy 4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db 10123 AAAGGGCAGGGTATGCCCCGCTCAATTTGGAGATCACTGTCATGTCCTCGGAGGTTTTGC 10182
Qy 4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db 10183 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCAAAAA 10242
Qy 5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db 10243 TCAAATGCTGCGGCTCCTTGGAATGTCAGCCGGCCGTTCATGCAGACTATACCTGCAAGG 10302
Qy 5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db 10303 TCTTCGGAGGGGTCTACCCCTTTATGTGGGGAGGAGCGCAATGTTTTTGCGACAGTGAGA 10362
Qy 5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db 10363 ACAGCCAGATGAGTGAGGCGTACGTCGAACTGTCAGCAGATTGCGCGTCTGACCACGCGC 10422
Qy 5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db 10423 AGGCGATTAAGGTGCACACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10482
Qy 5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db 10483 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACTTGA 10542
Qy 5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db 10543 AAGTCATAGCTGGACCAATTTCAGCATCGTTTACGCCATTCGATCATAAGGTCGTTATCC 10602
Qy 5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db 10603 ATCGCGGCCTGGTGTACAACTATGACTTCCCGGAATATGGAGCGATGAAACCAGGAGCGT 10662
Qy 5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db 10663 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAGGATCTCATCGCCAGCACAGACATTA 10722
Qy 5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db 10723 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCAGGAT 10782
Qy 5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db 10783 TTGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCACCTTTCGGGTGTA 10842
Qy 5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db 10843 AGATTGCAGTAAATCCGCTCCGAGCGGTGGACTGTTCATACGGGAACATTCCCATTTCTA 10902
Qy 5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db 10903 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10962
Qy 5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db 10963 AATGTGAAGTCAGTGAGTGCACTTATTCAGCAGACTTCGGCGGGATGGCCACCCTGCAGT 11022
Qy 5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db 11023 ATGTATCCGACCGCGAAGGTCAATGCCCCGTACATTCGCATTCGAGCACAGCAACTCTCC 11082
Qy 5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db 11083 AAGAGTCGACAGTACATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTTAGCACCGCGA 11142

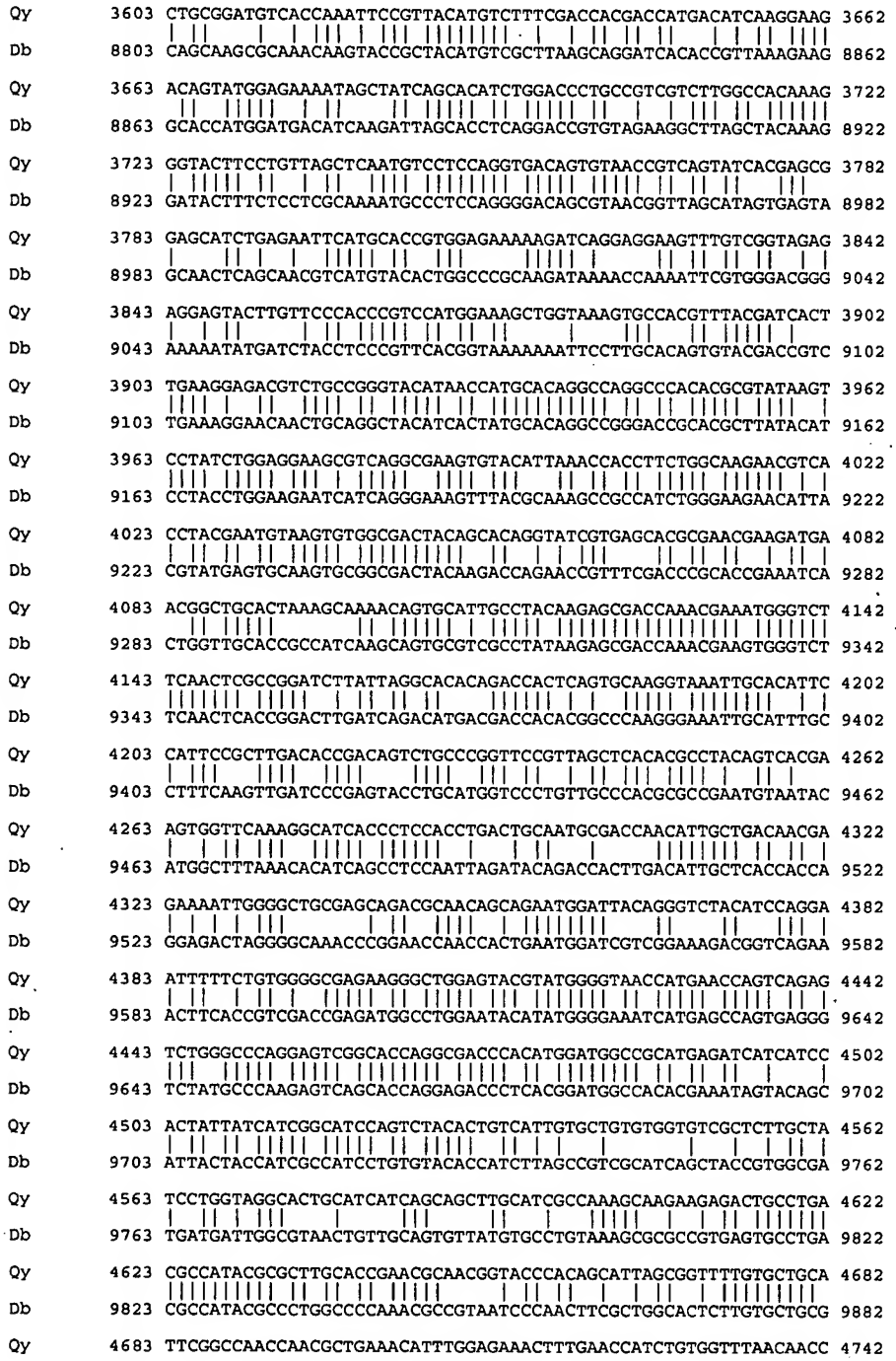
```
Qy       5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db      11143 GTCCACAGGCGAACTTTATCGTATCGCTGTGTGGGAAGAAGACAACATGCAATGCAGAAT 11202

Qy       6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db      11203 GTAAACCACCAGCTGACCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTTCAAG 11262

Qy       6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db      11263 CCGCCATCTCAAAAACATCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11322

Qy       6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
Db      11323 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11382

Qy       6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
Db      11383 GACCGCTACGCCCCAATGATCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11442

Qy       6242 GCATAATGCCACGCGCCGCT 6261
Db      11443 GCATAATGCATCAGGCTGGT 11462
```

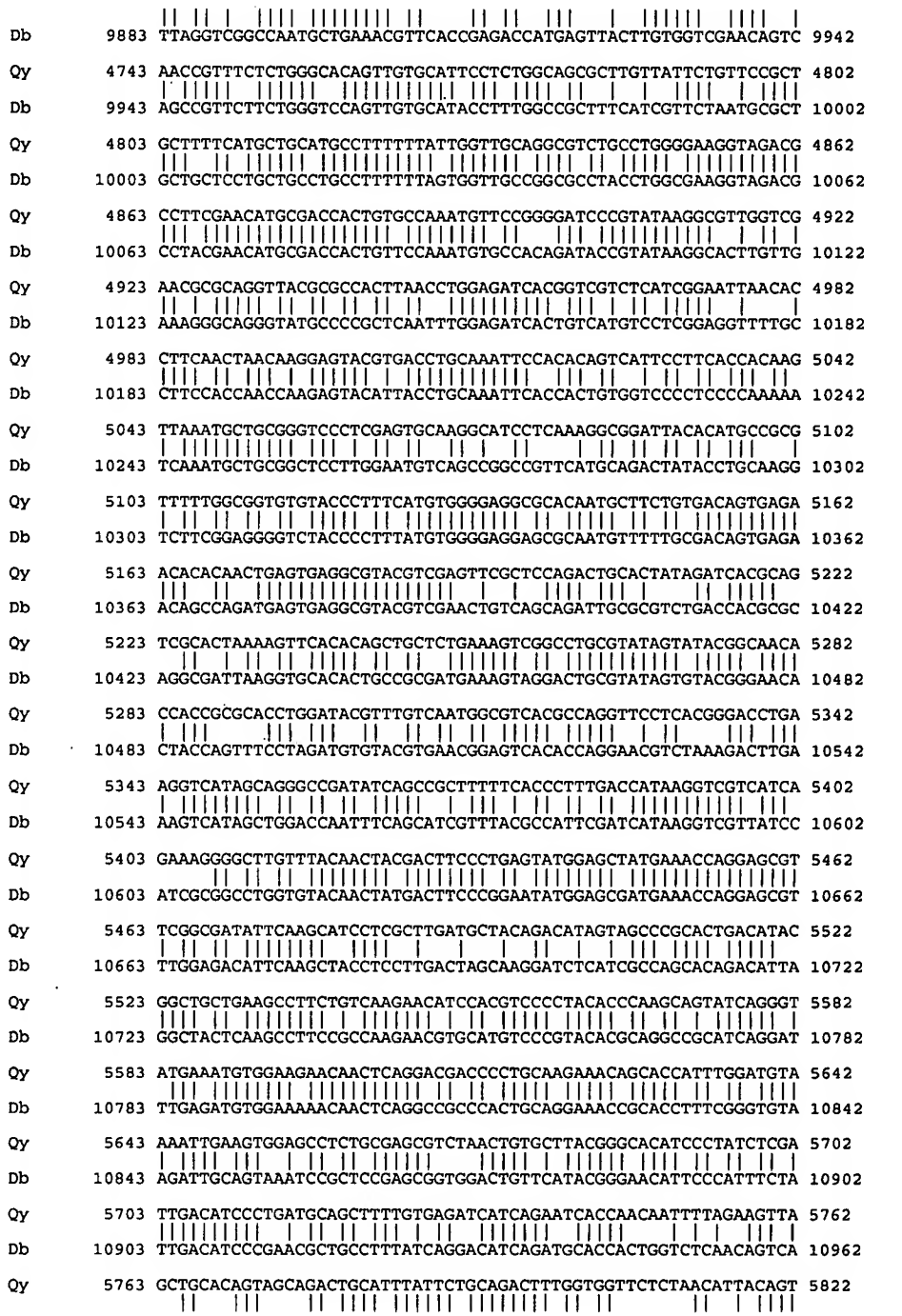
RESULT 7
AAC64507
ID AAC64507 standard; DNA; 11703 BP.
XX
AC AAC64507;
XX
DT 14-FEB-2001 (first entry)
XX
DE Alphavirus SinChironLP virus nucleotide sequence.
XX
KW Alphavirus; SinDCChiron virus; SinChironLP virus; immune response;
KW infection; human dendritic cell; immunostimulatory; cytostatic;
KW virucide; fungicide; antibacterial; antiparasitic; vaccine; cancer;
KW pathogen; antigen presenting cell; ds.
XX
OS Alphavirus.
XX
PN WO200061772-A2.
XX
PD 19-OCT-2000.
XX
PF 14-APR-2000; 2000WO-US10722.
XX
PR 14-APR-1999; 99US-0129498.
PR 09-AUG-1999; 99US-0148086.
PR 22-MAR-2000; 2000US-0191363.
XX
PA (CHIR) CHIRON CORP.
XX
PI Polo JM, Dubensky TW, Frolov I, Gardner JP, Otten G, Barnett S;
PI Driver DA;
XX
DR WPI; 2000-619231/59.
XX
PT New alphavirus that infects human dendritic cells for use in generating
PT an immune response to pathogenic agents such as bacteria, viruses,
PT fungi, parasites and cancer and for biological assays -
XX
PS Claim 10; Fig 2C; 83pp; English.
XX
CC The present invention describes an isolated alphavirus (AV) which
CC infects human dendritic cells and is not of American Type Culture
CC Collection (ATCC) number VR-2526. AAC64506 and AAC64507 represent the
CC nucleotide sequence of the specifically claimed SinDCChiron virus and
CC SinChironLP virus. The new AVs have immunostimulatory, cytostatic,
CC virucide, fungicide, antibacterial and antiparasitic activities and
CC can be used in vaccines. The AVs are used to infect dendritic cells,
CC preferably human cells. A heterologous sequence can be introduced and
CC expressed in human macrophages or antigen presenting cells in vivo and
CC in vitro, for use in biological assays. The AV-based vector systems
CC are used to generate an immune response to cancer or a pathogenic
CC agent, such as, bacteria, fungi, parasites or viruses. The AV can be
CC used to infect human dendritic cells, macrophages or antigen
CC presenting cells that previously could not be infected using an AV or
CC AV variant. The AV vectors are targeted directly to antigen presenting
CC cells.
XX
SQ Sequence 11703 BP; 3308 A; 3050 C; 2911 G; 2434 T; 0 other;

Query Match 36.7%; Score 2382; DB 21; Length 11703;
Best Local Similarity 64.6%; Pred. No. 0;
Matches 3694; Conservative 0; Mismatches 1985; Indels 41; Gaps 8;

```
Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
Db       5746 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5805

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
Db       5806 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTCCTGG 5865

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
Db       5866 AAAGAATTCATGCCCCGGTGCTCGACACGTCGAAAGAGGAACAACTCAAACTCAGGTACC 5925

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
Db       5926 AGATGATGCCCACCGAAGCCAACAAAAGTAGGTACCAGTCTCGTAAAGTAGAAAATCAGA 5985

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
Db       5986 AAGCCATAACCACTGAGCGACTACTGTCAGGACTACGACTGTATAACTCTGCCACAG--- 6042

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
Db       6043 ATCAGCCAGAATGCTATAAGATCACCTATCCGAAACCATTGTACTCCAGTAGCGTACCGG 6102

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
Db       6103 CGAACTACTCCGATCCACAGTTCGCTGTAGCTGTCTGTAACAACTATCTGCATGAGAACT 6162

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
Db       6163 ATCCGACAGTAGCATCTTATCAGATTACTGACGAGTACGATGCTTACTTGGATATGGTAG 6222

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
Db       6223 ACGGGACAGTCGCCTGCCTGGATACTGCAACCTTCTGCCCCGCTAAGCTTAGAAGTTACC 6282

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
Db       6283 CGAAAAAACATGAGTATAGAGCCCCGAATATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6342

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
Db       6343 ACACGCTACAAAATGTGCTCATTGCCGCAACTAAAAGAAATTGCAACGTCACGCAGATGC 6402

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
Db       6403 GTGAACTGCCAACACTGGACTCAGCGACATTCAATGTCGAATGCTTTCGAAAATATGCAT 6462

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
Db       6463 GTAATGACGAGTATTGGGAGGAGTTCGCTCGGAAGCCAATTAGGATTACCACTGAGTTTG 6522

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
Db       6523 TCACCGCATATGTAGCTAGACTGAAAGGCCCTAAGGCCGCCGCACTATTTGCAAAGACGT 6582

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
```


Db 6583 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGGG 6642
Qy 1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
Db 6643 ACGTGAAAGTTACACCAGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6702
Qy 1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
Db 6703 AAGCCGCAGAACCCCTGGCGACTGCTTACTTATGCGGGATTCACCGGGAATTAGTGCGTA 6762
Qy 1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
Db 6763 GGCTTACGGCCGTCTTGCTTCCAAACATTCACACGCTTTTTGACATGTCGGCGGAGGATT 6822
Qy 1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
Db 6823 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGCGACCCGGTACTGGAGACGGATATCG 6882
Qy 1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
Db 6883 CATCATTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGTCTGATGATCTTGGAGG 6942
Qy 1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
Db 6943 ACCTGGGTGTGGATCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 7002
Qy 1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
Db 7003 CCACCCATCTACCTACGGGTACTCGTTTTAAATTCGGGGCGATGATGAAATCCGGAATGT 7062
Qy 1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
Db 7063 TCCTCACACTTTTTGTCAACACAGTTTTGAATGTCGTTATCGCCAGCAGAGTACTAGAAG 7122
Qy 1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
Db 7123 AGCGGCTTAAAACGTCCAGATGTGCAGCGTTCATTGGCGACGACAACATCATACATGGAG 7182
Qy 2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
Db 7183 TAGTATCTGACAAAGAAATGGCTGAGAGGTGCGCCACCTGGCTCAACATGGAGGTTAAGA 7242
Qy 2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139
Db 7243 TCATCGACGCAGTCATCGGTGAGAGACCACCTTACTTCTGCGGCGGATTTATCTTGCAAG 7302
Qy 2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199
Db 7303 ATTCGGTTACTTCCACAGCGTGCCGCGTGGCGGACCCCCTGAAAAGGCTGTTTAAGTTGG 7362
Qy 2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
Db 7363 GTAAACCGCTCCCAGCCGACGACGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 7422
Qy 2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319
Db 7423 AAACAAAGGCGTGGTTTAGAGTAGGTATAACAGGCACTTTAGCAGTGGCCGTGACGACCC 7482
Qy 2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379
Db 7483 GGTATGAGGTAGACAATATTACACCTGTCCTACTGGCATTGAGAACTTTTGCCCAGAGCA 7542
Qy 2380 TTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGG 2439
Db 7543 AAAGAGCATTCCAAGCCATCAGAGGGGAAATAAAGCATCTCTACGGTGGTCCTAAATAGT 7602
Qy 2440 TGACGTAGTAGACACGCACCTAC----------CCACCGGCAGAATGTTTCCATACC 2486
Db 7603 CAGCATAGTACATTTCATCTGACTAATACTACAACACCACCACCATGAATAGAGGATTCT 7662
Qy 2487 CTCAGCTGAACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACC 2546
Db 7663 TTAACATGCTCGGCCGCCGCCCCTTCCCGGCCCCCACTGCCATGTGGAGGCCGCGGAGAA 7722
Qy 2547 CTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTA 2606
Db 7723 GGAGGCAGGCGGCCCCGATGCCTGCCCGCAACGGGCTGGCTTCTCAAATCCAGCAACTGA 7782
Qy 2607 GGAGGTCGATAGTCAACTTGACTTTCAAACAACGATCACCTAATCCGCCGCCAGGTCCA- 2665
Db 7783 CCACAGCCGTCAGTGCCCTAGTCATTGGACAGGCAACTAGACCTCAACCCCCACGTCCAC 7842
Qy 2666 -----CCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGA 2720
Db 7843 GCCCGCCACCGCGCCAGAAGAAGCAGGCGCCCAAGCAACCACCGAAGCCGAAGAAACCAA 7902
Qy 2721 AGAAGCAGCAAGCCAAGAGGACGAAAC---GCAAGCCTAAACCAGGGAAACGACAACGTA 2777
Db 7903 AAACGCAGGAGAAGAAGAAGAAGCAACCTGCAAAACCCAAACCCGGAAAGAGACAGCGCA 7962
Qy 2778 TGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATCATG---CTGAACGGCCAAGTGA 2834
Db 7963 TGGCACTTAAGTTGGAGGCCGACAGATCGTTCGACGTCAAGAACGAGGACGGAGATGTCA 8022
Qy 2835 ATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACTCCACGTTGAAGGAAAAA 2894
Db 8023 TCGGGCACGCACTGGCCATGGAAGGAAAGGTAATGAAACCTCTGCACGTGAAAGGAACCA 8082
Qy 2895 TTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAGCATGTACGACTTGGAGT 2954
Db 8083 TCGACCACCCTGTGCTATCAAAGCTCAAATTTACCAAGTCGTCAGCATACGACATGGAGT 8142
Qy 2955 ACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTACACCAGCGACAAACCAC 3014
Db 8143 TCGCACAGTTGCCAGTCAACATGAGAAGTGAGGCATTCACCTACACCAGTGAACACCCCG 8202
Qy 3015 CGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAATGGGAGATTTACCGTAC 3074
Db 8203 AAGGATTCTATAACTGGCACCACGGAGCGGTGCAGTATAGTGGAGGTAGATTTACCATCC 8262
Qy 3075 CGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGACAACAGAGGCAGAG 3134
Db 8263 CTCGCGGAGTAGGAGGCAGAGGAGACAGCGGTCGTCCGATCATGGATAACTCCGGTCGGG 8322
Qy 3135 TTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGCGCTTTCAGTGGTCA 3194
Db 8323 TTGTCGCGATAGTCCTCGGTGGAGCTGATGAAGGAACACGAACTGCCCTTTCGGTCGTCA 8382
Qy 3195 CTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTCTGAACCGTGGT--- 3251
Db 8383 CCTGGAATAGTAAAGGGAAGACAATTAAGACGACCCCGGAAGGGACAGAAGAGTGGTCCG 8442
Qy 3252 ------CACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGACAAAC 3305
Db 8443 CAGCACCACTGGTCACGGCAATGTGTTTGCTCGGAAATGTGAGCTTCCCATGCGACCGCC 8502
Qy 3306 CACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAACGTCG 3365
Db 8503 CGCCCACATGCTATACCCGCGAACCTTCCAGAGCCCTCGACATCCTTGAAGAGAACGTGA 8562
Qy 3366 ACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAATGTCCATCAC---GCCGGC 3422
Db 8563 ACCATGAGGCCTACGATACCCTGCTCAATGCCATATTGCGGTGCGGATCGTCTGGCAGAA 8622
Qy 3423 CCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCCCGT 3482
Db 8623 GCAAAAGAAGCGTCACTGACGACTTTACCCTGACCAGCCCCTACTTGGGCACATGCTCGT 8682
Qy 3483 ATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACGAAT 3542
Db 8683 ACTGCCACCATACTGAACCGTGCTTCAGCCCTGTTAAGATCGAGCAGGTCTGGGACGAAG 8742
Qy 3543 CTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAGGCA 3602
Db 8743 CGGACGATAACACCATACGCATACAGACTTCCGCCCAGTTTGGATACGACCAAAGCGGAG 8802



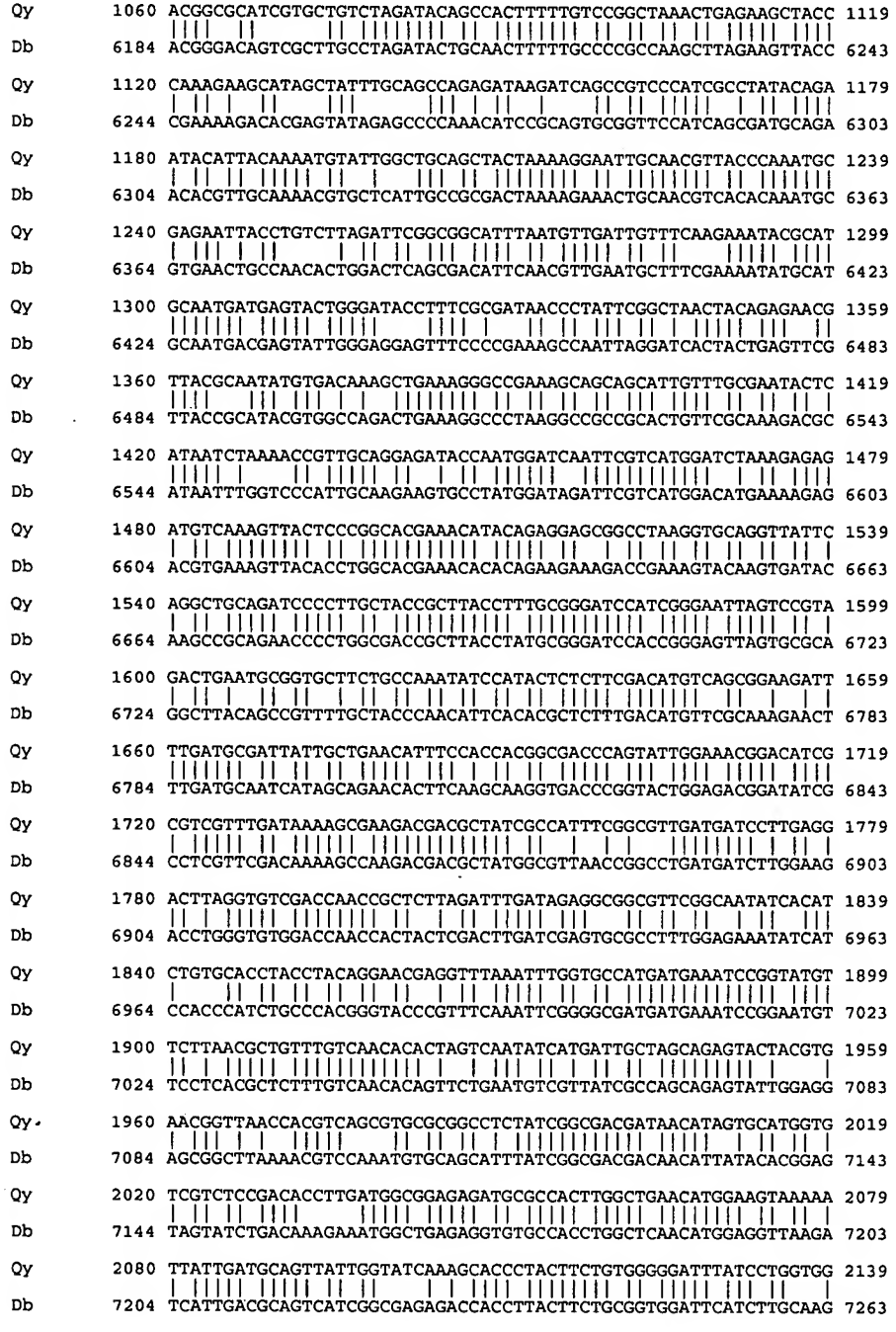


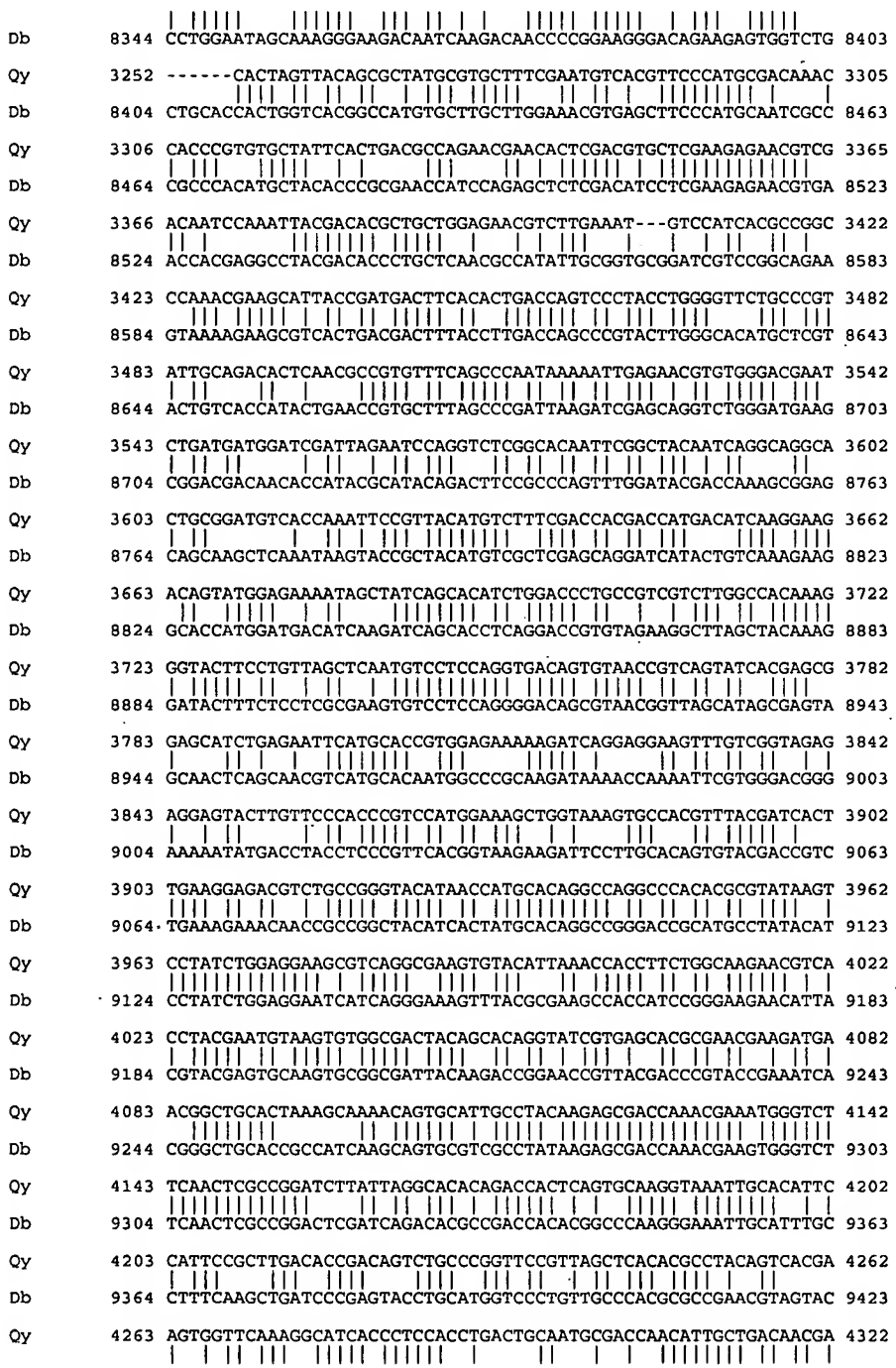
```
Db      10963 AATGTGAAGTCAGTGAGTGCACTTATTCAGCAGACTTCGGCGGGATGGCCACCCTGCAGT 11022

Qy       5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882

Db      11023 ATGTATCCGACCGCGAAGGTCAATGCCCCGTACATTCGCATTCGAGCACAGCAACTCTCC 11082

Qy       5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942

Db      11083 AAGAGTCGACAGTACATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTTAGCACCGCGA 11142

Qy       5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002

Db      11143 GTCCACAGGCGAACTTTATCGTATCGCTGTGTGGGAAGAAGACAACATGCAATGCAGAAT 11202

Qy       6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062

Db      11203 GTAAACCACCAGCTGACCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTTCAAG 11262

Qy       6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122

Db      11263 CCGCCATCTCAAAAACATCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11322

Qy       6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182

Db      11323 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11382

Qy       6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241

Db      11383 GACCGCTACGCCCCAATGATCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11442

Qy       6242 GCATAATGCCACGCGCCGCT 6261

Db      11443 GCATAATGCATCAGGCTGGT 11462
```

```
RESULT 8
AAT47668
ID   AAT47668 standard; cDNA; 11663 BP.
XX
AC   AAT47668;
XX
DT   19-MAY-1997  (first entry)
XX
DE   South African Arbovirus 86 RNA transcript cDNA.
XX
KW   S.A.AR86; Sindbis virus; attenuation; vaccine; diagnosis; antibody;
KW   vector; Ockelbo; polyarthritis; ss.
XX
OS   South African Arbovirus No. 86.
XX
FH   Key             Location/Qualifiers
FT   mutation        215
FT                   /*tag=  a
FT                   /note= "a guanine for adenine substn. at nucleotide
FT                          215 is a preferred silent mutation used in
FT                          recombinant DNA constructs of the
FT                          invention"
FT   mutation        3863
FT                   /*tag=  b
FT                   /note= "a guanine for cytosine substn. at
FT                          nucleotide 3863 is a preferred silent
FT                          mutation used in recombinant DNA constructs
FT                          of the invention"
FT   mutation        5984
FT                   /*tag=  c
FT                   /note= "a guanine for adenine substn. at nucleotide
FT                          5984 is a preferred silent mutation used in
FT                          recombinant DNA constructs of the
FT                          invention"
FT   mutation        9113
FT                   /*tag=  d
FT                   /note= "a cytosine for thymine substn. at
FT                          nucleotide 9113 is a preferred silent
FT                          mutation used in recombinant DNA constructs
FT                          of the invention"
XX
PN   WO9637220-A1.
XX
PD   28-NOV-1996.
XX
PF   22-MAY-1996;   96WO-US07457.
XX
PR   23-MAY-1995;   95US-0446932.
XX
PA   (UYNC-) UNIV NORTH CAROLINA.
XX
PI   Davis NL,  Johnston RE,  Simpson DA;
XX
DR   WPI; 1997-020937/02.
XX
PT   New recombinant DNA encoding infectious South African Arbovirus 86
PT   RNA transcript and related mutants - under control of heterologous
PT   promoter, useful in prodn. of attenuated vaccines and to generate
PT   diagnostic antibodies
XX
PS   Claim 2; Page 13-19; 27pp; English.
XX
CC   A cDNA clone (AAT47668) codes for an infectious South African
CC   Arbovirus No. 86 (S.A.AR86) virus RNA transcript.  Novel
CC   recombinant DNA constructs comprise the cDNA, pref. mutated to
CC   include attenuating and silent mutations, and an upstream
CC   heterologous promoter (pref. T3, T7 or SP6).  Infectious
CC   attenuated viral particles produced from cells transfected with the
CC   RNA transcripts encoded by such cDNA clones are useful in live
CC   attenuated vaccines, e.g. against diseases caused by related
CC   viruses such as Ockelbo, and can also be used as vectors to express
CC   antigens of other viruses.  The cDNA clone can be administered to
CC   animals to raise antibodies for diagnosis of S.A.AR86 infection.
XX
SQ   Sequence 11663 BP; 3297 A; 3104 C; 2903 G; 2359 T; 0 other;
```

Query Match 36.7%; Score 2381.6; DB 18; Length 11663;
Best Local Similarity 64.4%; Pred. No. 0;
Matches 3742; Conservative 0; Mismatches 1994; Indels 74; Gaps 9;

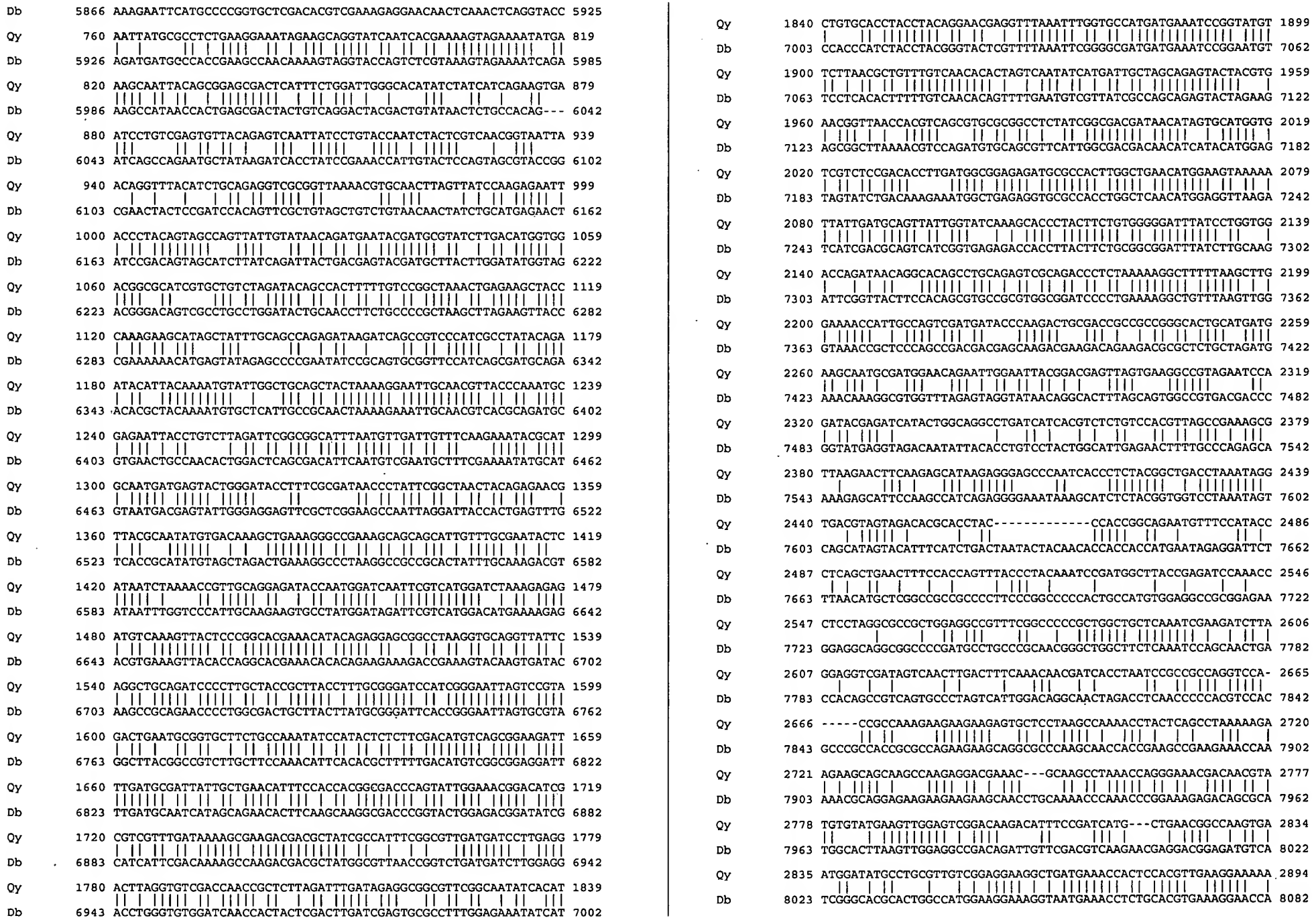
```
Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639

Db       5707 ACTGTCTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5766

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699

Db       5767 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTTCTGG 5826

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759

Db       5827 AAAGAATCTACGCCCCGGTGCTCGACACGTCGAAAGAGGAACAGCTCAAACTCAGGTACC 5886

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819

Db       5887 AGATGATGCCCACCGGAACCAACAAAAGCAGGTACCAGTCTCGAAAAGTAGAAAACCAGA 5946

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879

Db       5947 AAGCCATAACCACTGAGCGACTGCTTTCAGGGCTACGACTGTATAACTCTGCCACAG--- 6003

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939

Db       6004 ATCAGCCAGAATGCTATAAGATCACCTACCCGAAACCATCGTATTCCAGCAGTGTACCAG 6063.

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999

Db       6064 CGAACTACTCTGACCCAAAGTTTGCTGTAGCTGTTTGTAACAACTATCTGCATGAGAATT 6123

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059

Db       6124 ACCCGACGGTAGCATCTTATCAGATCACCGACGAGTACGATGCTTACTTGGATATGGTAG 6183
```

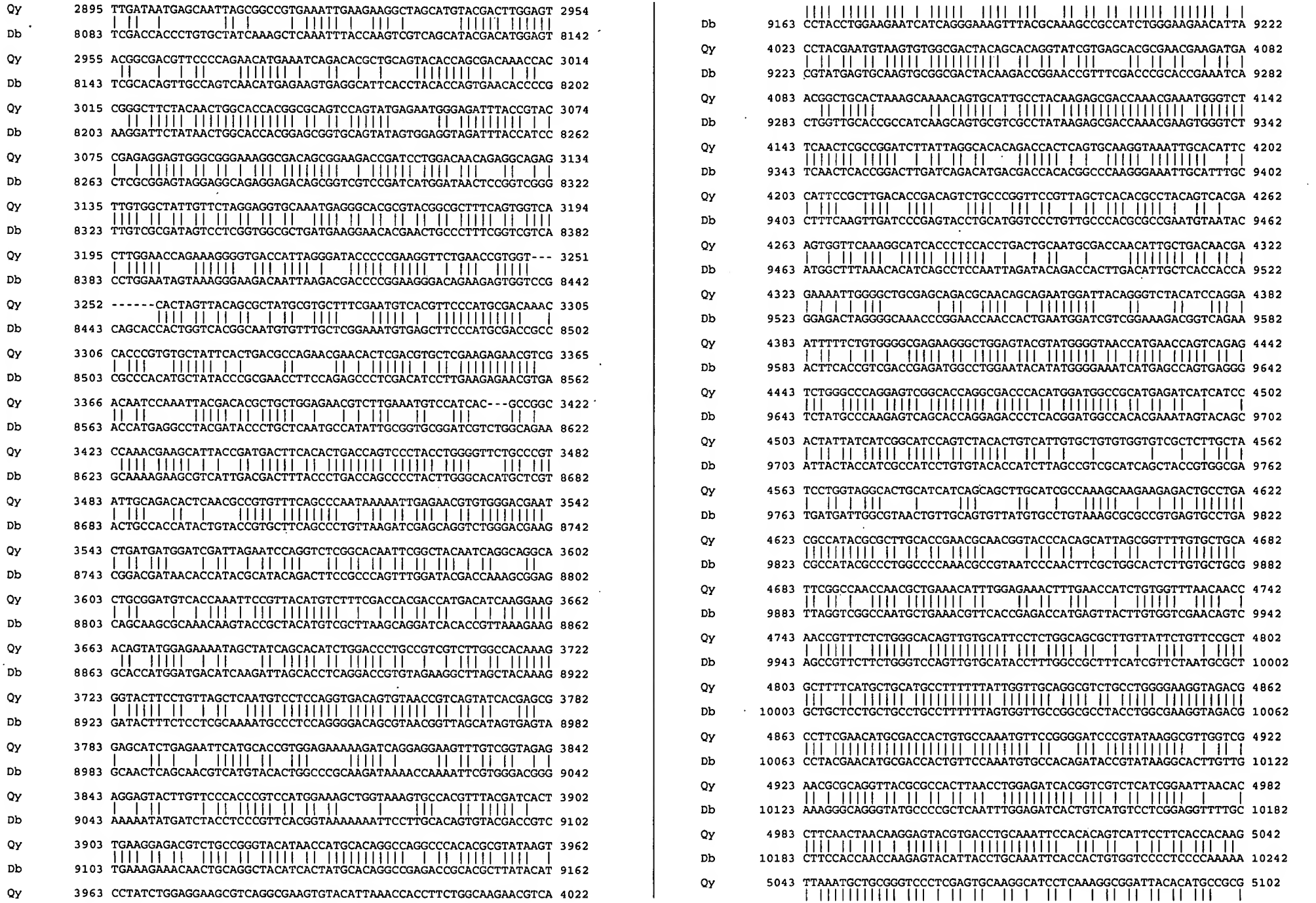

Qy 1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
Db 6184 ACGGGACAGTCGCTTGCCTAGATACTGCAACTTTTTGCCCCGCCAAGCTTAGAAGTTACC 6243
Qy 1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
Db 6244 CGAAAAGACACGAGTATAGAGCCCCAAACATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6303
Qy 1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
Db 6304 ACACGTTGCAAAACGTGCTCATTGCCGCGACTAAAAGAAACTGCAACGTCACACAAATGC 6363
Qy 1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
Db 6364 GTGAACTGCCAACACTGGACTCAGCGACATTCAACGTTGAATGCTTTCGAAAATATGCAT 6423
Qy 1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
Db 6424 GCAATGACGAGTATTGGGAGGAGTTTCCCCGAAAGCCAATTAGGATCACTACTGAGTTCG 6483
Qy 1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
Db 6484 TTACCGCATACGTGGCCAGACTGAAAGGCCCTAAGGCCGCCGCACTGTTCGCAAAGACGC 6543
Qy 1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
Db 6544 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6603
Qy 1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
Db 6604 ACGTGAAAGTTACACCTGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6663
Qy 1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
Db 6664 AAGCCGCAGAACCCCTGGCGACCGCTTACCTATGCGGGATCCACCGGGAGTTAGTGCGCA 6723
Qy 1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
Db 6724 GGCTTACAGCCGTTTTGCTACCCAACATTCACACGCTCTTTGACATGTTCGCAAAGAACT 6783
Qy 1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
Db 6784 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGTGACCCGGTACTGGAGACGGATATCG 6843
Qy 1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
Db 6844 CCTCGTTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGCCTGATGATCTTGGAAG 6903
Qy 1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
Db 6904 ACCTGGGTGTGGACCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 6963
Qy 1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
Db 6964 CCACCCATCTGCCCACGGGTACCCGTTTCAAATTCGGGGCGATGATGAAATCCGGAATGT 7023
Qy 1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
Db 7024 TCCTCACGCTCTTTGTCAACACAGTTCTGAATGTCGTTATCGCCAGCAGAGTATTGGAGG 7083
Qy 1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
Db 7084 AGCGGCTTAAAACGTCCAAATGTGCAGCATTTATCGGCGACGACAACATTATACACGGAG 7143
Qy 2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
Db 7144 TAGTATCTGACAAAGAAATGGCTGAGAGGTGTGCCACCTGGCTCAACATGGAGGTTAAGA 7203
Qy 2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139
Db 7204 TCATTGACGCAGTCATCGGCGAGAGACCACCTTACTTCTGCGGTGGATTCATCTTGCAAG 7263
Qy 2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199
Db 7264 ATTCGGTTACCTCCACAGCGTGTCGCGTGGCGGACCCCTTGAAAAGGCTGTTTAAGTTGG 7323
Qy 2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
Db 7324 GTAAACCGCTCCCAGCCGACGATGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 7383
Qy 2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319
Db 7384 AAACAAAGGCGTGGTTTAGAGTAGGTATACAGACACACTTAGCAGTGGCCGTGGCAACTC 7443
Qy 2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379
Db 7444 GGTATGAGGTAGACAACATCACACCTGTCCTGCTGGCATTGAGAACTTTTGCCCAGAGCA 7503
Qy 2380 TTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGG 2439
Db 7504 AAAGAGCATTTCAAGCCATCAGAGGGGAAATAAAGCATCTCTACGGTGGTCCTAAATAGT 7563
Qy 2440 TGACGTAGTAGA----------CACGCACCTACCCACCGGCAGAATGTTTCCA 2482
Db 7564 CAGCATAGTACATTTCATCTGACTAATACCACAACACCACCACCATGAATAGAGGATTCT 7623
Qy 2483 TACCCTCAGCTGAACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCA 2542
Db 7624 TTAACATGCTCGGCCGCCGCCCCTTCCCAGCCCCCACTGCCATGTGGAGGCCGCGGAGAA 7683
Qy 2543 AACCCTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGAT 2602
Db 7684 GGAGGCAGGCGGCCCCGATGCCTGCCCGCAATGGGCTGGCTTCCCAAATCCAGCAACTGA 7743
Qy 2603 CTTAGGAGGTCGATAGTCAACTTGACTTTCAAACAAC--GATCACCTAATCCGCCGCCAG 2660
Db 7744 CCACAGCCGTCAGTGCCCTAGTCATTGGACAGGCAACTAGACCTCAAACCCCACGCCCAC 7803
Qy 2661 GTCCACCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGA 2720
Db 7804 GCCCGCCGCCGCGCCAGAAGAAGCAGGCGCCAAAGCAACCACCGAAGCCGAAGAAACCAA 7863
Qy 2721 AGAAGCAGCAAGCCAAGAGGACGAAAC---GCAAGCCTAAACCAGGGAAACGACAACGTA 2777
Db 7864 AAACACAGGAGAAGAAGAAGAAGCAACCTGCAAAACCCAAACCCGGAAAGAGACAGCGTA 7923
Qy 2778 TGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATC---ATGCTGAACGGCCAAGTGA 2834
Db 7924 TGGCACTTAAGTTGGAGGCCGACAGACTGTTCGACGTCAAAAATGAGGACGGAGATGTCA 7983
Qy 2835 ATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACTCCACGTTGAAGGAAAAA 2894
Db 7984 TCGGGCACGCACTGGCCATGGAAGGAAAGGTAATGAAACCACTCCACGTGAAAGGAACTA 8043
Qy 2895 TTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAGCATGTACGACTTGGAGT 2954
Db 8044 TTGACCACCCTGTGCTATCAAAGCTCAAATTCACCAAGTCGTCAGCATACGACATGGAGT 8103
Qy 2955 ACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTACACCAGCGACAAACCAC 3014
Db 8104 TCGCACAGTTGCCGGTCAACATGAGAAGTGAGGCGTTCACCTACACCAGTGAAGACCCTG 8163
Qy 3015 CGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAATGGGAGATTTACCGTAC 3074
Db 8164 AAGGGTTCTACAACTGGCACCACGGAGCGGTGCAGTATAGTGGAGGCAGATTTACCATCC 8223
Qy 3075 CGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGACAACAGAGGCAGAG 3134
Db 8224 CCCGCGGAGTAGGAGGCAGAGGAGACAGTGGTCGTCCGATTATGGATAACTCAGGCCGGG 8283
Qy 3135 TTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGCGCTTTCAGTGGTCA 3194
Db 8284 TTGTCGCGATAGTCCTCGGAGGGGCTGATGAGGGAACAAGAACCGCCCTTTCGGTCGTCA 8343
Qy 3195 CTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTCTGAACCGTGGT--- 3251


Db 8344 CCTGGAATAGCAAAGGGAAGACAATCAAGACAACCCCGGAAGGGACAGAAGAGTGGTCTG 8403
Qy 3252 ------CACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGACAAAC 3305
Db 8404 CTGCACCACTGGTCACGGCCATGTGCTTGCTTGGAAACGTGAGCTTCCCATGCAATCGCC 8463
Qy 3306 CACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAACGTCG 3365
Db 8464 CGCCCACATGCTACACCCGCGAACCATCCAGAGCTCTCGACATCCTCGAAGAGAACGTGA 8523
Qy 3366 ACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAAT---GTCCATCACGCCGGC 3422
Db 8524 ACCACGAGGCCTACGACACCCTGCTCAACGCCATATTGCGGTGCGGATCGTCCGGCAGAA 8583
Qy 3423 CCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCCCGT 3482
Db 8584 GTAAAAGAAGCGTCACTGACGACTTTACCTTGACCAGCCCGTACTTGGGCACATGCTCGT 8643
Qy 3483 ATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACGAAT 3542
Db 8644 ACTGTCACCATACTGAACCGTGCTTTAGCCCGATTAAGATCGAGCAGGTCTGGGATGAAG 8703
Qy 3543 CTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAGGCA 3602
Db 8704 CGGACGACAACACCATACGCATACAGACTTCCGCCCAGTTTGGATACGACCAAAGCGGAG 8763
Qy 3603 CTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAAGGAAG 3662
Db 8764 CAGCAAGCTCAAATAAGTACCGCTACATGTCGCTCGAGCAGGATCATACTGTCAAAGAAG 8823
Qy 3663 ACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCACAAAG 3722
Db 8824 GCACCATGGATGACATCAAGATCAGCACCTCAGGACCGTGTAGAAGGCTTAGCTACAAAG 8883
Qy 3723 GGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCACGAGCG 3782
Db 8884 GATACTTTCTCCTCGCGAAGTGTCCTCCAGGGGACAGCGTAACGGTTAGCATAGCGAGTA 8943
Qy 3783 GAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTAGAG 3842
Db 8944 GCAACTCAGCAACGTCATGCACAATGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGG 9003
Qy 3843 AGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATCACT 3902
Db 9004 AAAAATATGACCTACCTCCCGTTCACGGTAAGAAGATTCCTTGCACAGTGTACGACCGTC 9063
Qy 3903 TGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAAGT 3962
Db 9064 TGAAAGAAACAACCGCCGGCTACATCACTATGCACAGGCCGGGACCGCATGCCTATACAT 9123
Qy 3963 CCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGTCA 4022
Db 9124 CCTATCTGGAGGAATCATCAGGGAAAGTTTACGCGAAGCCACCATCCGGGAAGAACATTA 9183
Qy 4023 CCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGATGA 4082
Db 9184 CGTACGAGTGCAAGTGCGGCGATTACAAGACCGGAACCGTTACGACCCGTACCGAAATCA 9243
Qy 4083 ACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGTCT 4142
Db 9244 CGGGCTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAACGAAGTGGGTCT 9303
Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9304 TCAACTCGCCGGACTCGATCAGACACGCCGACCACACGGCCCAAGGGAAATTGCATTTGC 9363
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9364 CTTTCAAGCTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAACGTAGTAC 9423
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9424 ACGGCTTTAAACACATCAGCCTCCAATTAGACACAGACCATCTGACATTGCTCACCACCA 9483
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9484 GGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCATCGGAAACACGGTTAGAA 9543
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9544 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGCAATCACGAACCAGTAAGGG 9603
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db 9604 TCTATGCCCAAGAGTCTGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9663
Qy 4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db 9664 ATTACTATCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTGCTGTGGCGA 9723
Qy 4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db 9724 TGATGATTGGCGTAACTGTTGCAGCATTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9783
Qy 4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db 9784 CGCCATATGCCCTGGCCCCAAATGCCGTGATTCCAACTTCGCTGGCACTTTTGTGCTGTG 9843
Qy 4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db 9844 TTAGGTCGGCTAATGCTGAAACATTCACCGAGACCATGAGTTACTTATGGTCGAACAGCC 9903
Qy 4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
Db 9904 AGCCGTTCTTCTGGGTCCAGCTGTGTATACCTCTGGCCGCTGTCGTCGTTCTAATGCGCT 9963
Qy 4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
Db 9964 GTTGCTCATGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10023
Qy 4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
Db 10024 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10083
Qy 4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db 10084 AAAGGGCAGGGTACGCCCCGCTCAATTTGGAGATTACTGTCATGTCCTCGGAGGTTTTGC 10143
Qy 4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db 10144 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCTAAAG 10203
Qy 5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db 10204 TCAGATGCTGCGGCTCCTTGGAATGTCAGCCCGCCGCTCACGCAGACTATACCTGCAAGG 10263
Qy 5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db 10264 TCTTTGGAGGGGTGTACCCCTTCATGTGGGGAGGAGCACAATGTTTTTGCGACAGTGAGA 10323
Qy 5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db 10324 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGTAGATTGCGCGACTGACCACGCGC 10383
Qy 5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db 10384 AGGCGATTAAGGTGCATACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10443
Qy 5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db 10444 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACCTGA 10503
Qy 5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db 10504 AAGTCATAGCTGGACCAATTTCAGCATTGTTTACACCATTCGATCACAAGGTCGTTATCA 10563

```
Qy       5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db      10564 ATCGCGGCCTGGTGTACAACTATGACTTTCCGGAATACGGAGCGATGAAACCAGGAGCGT 10623

Qy       5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db      10624 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAAGACCTCATCGCCAGCACAGACATTA 10683

Qy       5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db      10684 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCTGGAT 10743

Qy       5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db      10744 TCGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAAACGCCCCTTTTGGGTGCA 10803

Qy       5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db      10804 AGATTGCAGTCAATCCGCTTCGAGCGGTGGACTGCTCATACGGGAACATTCCCATTTCTA 10863

Qy       5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db      10864 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10923

Qy       5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db      10924 AATGTGATGTCAGTGAGTGCACTTATTCAGCGGACTTCGGAGGGATGGCTACCCTGCAGT 10983

Qy       5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db      10984 ATGTATCCGACCGCGAAGGACAATGCCCTGTACATTCGCATTCGAGCACAGCAACCCTCC 11043

Qy       5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db      11044 AAGAGTCGACAGTTCATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTCAGCACCGCGA 11103

Qy       5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db      11104 GCCCACAGGCGAACTTCATTGTATCGCTGTGTGGTAAGAAGACAACATGCAATGCAGAAT 11163

Qy       6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db      11164 GCAAACCACCAGCTGATCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTCCAAG 11223

Qy       6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db      11224 CCGCCATCTCAAAAACTTCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11283

Qy       6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
Db      11284 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11343

Qy       6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
Db      11344 GACCGCTACGCCCCAATGACCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11403

Qy       6242 GCATAATGCCACGCGCCGCT----------------------------------TGACACT 6268
Db      11404 GCATAATGCATCAGGCTGGTATATTAGATCCCCGCTTACCGCGGGCAATATAGCAACACC 11463

Qy       6269 AAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGT 6318
Db      11464 AAAACTCGACGTATTTCCGAGGAAGCGCAGTGCATAATGCTGCGCAGTGT 11513
```

```
RESULT 9
ABX81526
ID   ABX81526 standard; cDNA; 11740 BP.
XX
AC   ABX81526;
XX
DT   23-APR-2003  (first entry)
XX
DE   Sindbis virus cDNA #3.
XX
KW   Eukaryotic layered vector initiation system; cytopathic effect; gene; ss;
KW   alphaviral nonstructural protein 2; RNA vector replicon; alphavirus RNA;
KW   macromolecular synthesis; alphavirus RNA polymerase recognition sequence;
KW   3' polyadenylate tract; cellular macromolecular synthesis; cell death;
KW   Sindbis virus.
XX
OS   Sindbis virus.
XX
PN   US6465634-B1.
XX
PD   15-OCT-2002.
XX
PF   08-OCT-1999;   99US-0415900.
XX
PR   06-OCT-1997;   97US-0944645.
PR   05-APR-1996;   96US-0628594.
PR   24-JUN-1996;   96US-0668953.
PR   12-JUL-1996;   96US-0679640.
PR   04-APR-1997;   97US-0833148.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Dubensky TW,  Polo JM,  Belli BA,  Schlesinger S,  Dryga SA,  Frolov I;
XX
DR   WPI; 2003-147073/14.
XX
PT   Eukaryotic layered vector initiation system, for gene therapy, has
PT   alphaviral nonstructural protein gene having mutant nonstructural
PT   protein 2 gene, which reduces host-cell directed macromolecular
PT   synthesis -
XX
PS   Example 1; Fig 8A-8E; 161pp; English.
XX
CC   The invention relates to a eukaryotic layered vector initiation system,
CC   comprising a nucleic acid sequence encoding all four alphaviral
CC   nonstructural proteins and including an altered sequence encoding for
CC   nonstructural protein 2, such that when the altered sequence is operably
CC   incorporated into an RNA vector replicon, the time required to reach 50%
CC   inhibition of cellular macromolecular synthesis in cells is increased, as
CC   compared to an RNA vector replicon having a wild-type alphavirus
CC   nonstructural protein 2. The initiation system comprises a 5' promoter
CC   which directs synthesis of alphavirus RNA in vivo from cDNA, a 5'
CC   sequence which directs transcription of alphavirus RNA, a nucleic acid
CC   sequence which operably encodes all four alphaviral nonstructural
CC   proteins, an alphavirus RNA polymerase recognition sequence and a 3'
CC   polyadenylate tract. The eukaryotic layered vector initiation system is
CC   useful for stimulating an immune response within a vertebrate, for
CC   protein expression and gene therapy. The system exhibits reduced, delayed
CC   or no inhibition of cellular macromolecular synthesis, thus permitting
CC   its use for protein expression and gene therapy with reduced, delayed or
CC   no development of cytopathic effects or cell death. This sequence
CC   represents Sindbis virus cDNA, used in the scope of the invention.
XX
SQ   Sequence 11740 BP; 3345 A; 3049 C; 2908 G; 2438 T; 0 other;

  Query Match           36.7%;  Score 2380.4;  DB 25;  Length 11740;
  Best Local Similarity   64.6%;  Pred. No. 0;
  Matches 3693;  Conservative    0;  Mismatches 1986;  Indels   41;  Gaps    8;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
Db       5746 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5805

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
Db       5806 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTCCTGG 5865

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
```





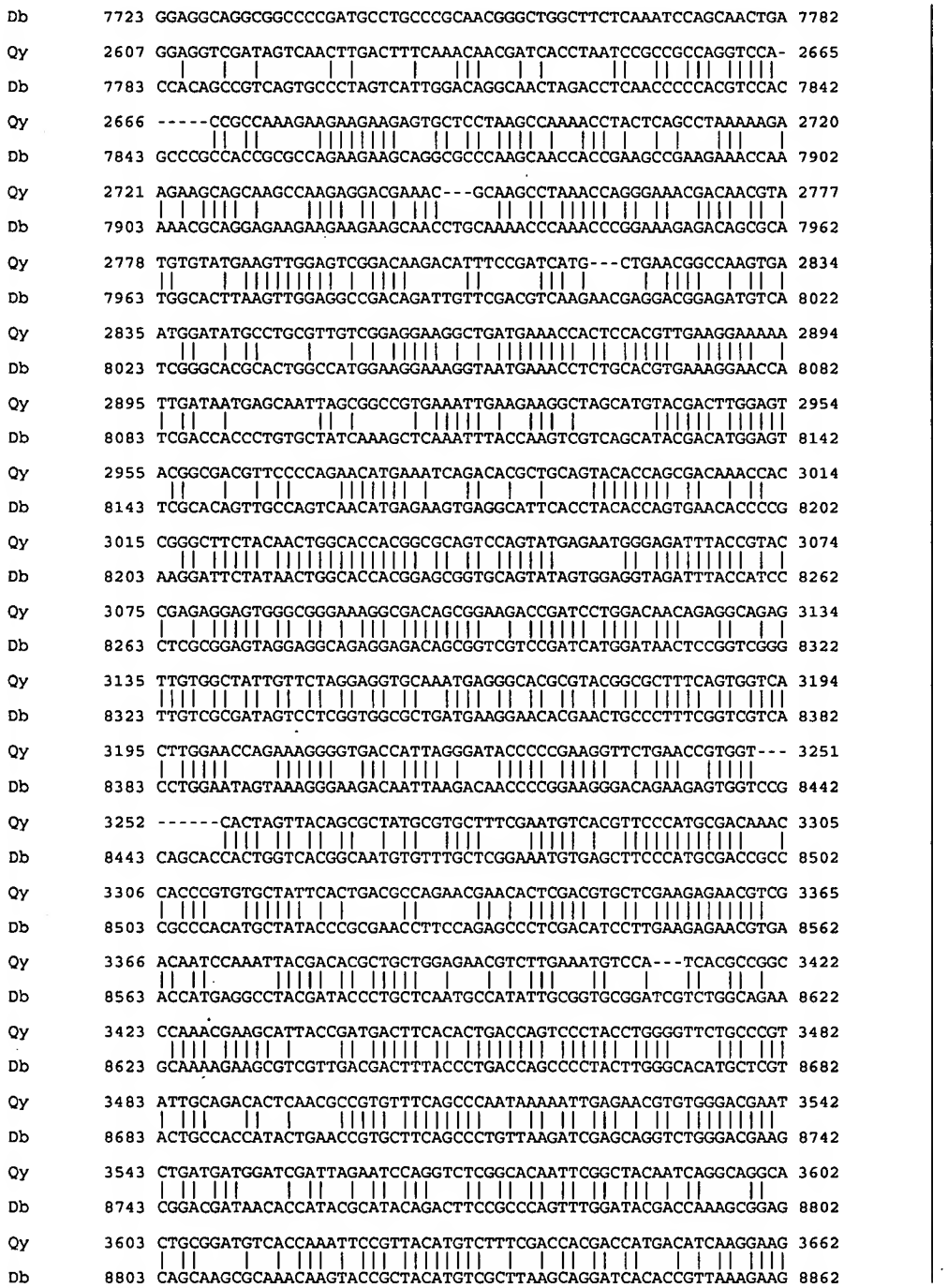
```
Db      10243 TCAAATGCTGCGGCTCCTTGGAATGTCAGCCGGCCGCTCATGCAGACTATACCTGCAAGG 10302

Qy       5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162

Db      10303 TCTTCGGAGGGGTCTACCCCTTTATGTGGGGAGGAGCGCAATGTTTTTGCGACAGTGAGA 10362

Qy       5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222

Db      10363 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGCAGATTGCGCGTCTGACCACGCGC 10422

Qy       5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282

Db      10423 AGGCGATTAAGGTGCACACTGCCGCGATGAAAGTAGGACTGCGTATTGTGTACGGGAACA 10482

Qy       5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342

Db      10483 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACTTGA 10542

Qy       5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402

Db      10543 AAGTCATAGCTGGACCAATTTCAGCATCGTTTACGCCATTCGATCATAAGGTCGTTATCC 10602

Qy       5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462

Db      10603 ATCGCGGCCTGGTGTACAACTATGACTTCCCGGAATATGGAGCGATGAAACCAGGAGCGT 10662

Qy       5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522

Db      10663 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAGGATCTCATCGCCAGCACAGACATTA 10722

Qy       5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582

Db      10723 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCTCATCAGGAT 10782

Qy       5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642

Db      10783 TTGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCACCTTTCGGGTGTA 10842

Qy       5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702

Db      10843 AGATTGCAGTAAATCCGCTCCGAGCGGTGGACTGTTCATACGGGAACATTCCCATTTCTA 10902

Qy       5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762

Db      10903 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10962

Qy       5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822

Db      10963 AATGTGAAGTCAGTGAGTGCACTTATTCAGCAGACTTCGGCGGGATGGCCACCCTGCAGT 11022

Qy       5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882

Db      11023 ATGTATCCGACCGCGAAGGTCAATGCCCCGTACATTCGCATTCGAGCACAGCAACTCTCC 11082

Qy       5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942

Db      11083 AAGAGTCGACAGTACATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTTAGCACCGCGA 11142

Qy       5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002

Db      11143 GTCCACAGGCGAACTTTATCGTATCGCTGTGTGGGAAGAAGACAACATGCAATGCAGAAT 11202

Qy       6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062

Db      11203 GTAAACCACCAGCTGACCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTTCAAG 11262

Qy       6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122

Db      11263 CCGCCATCTCAAAAACATCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11322

Qy       6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182

Db      11323 TATTAATTATAGGACTTATGATTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11382

Qy       6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241

Db      11383 GACCGCTACGCCCCAATGATCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11442

Qy       6242 GCATAATGCCACGCGCCGCT 6261

Db      11443 GCATAATGCATCAGGCTGGT 11462
```

RESULT 10
AAT30787
ID AAT30787 standard; DNA; 16656 BP.
XX
AC AAT30787;
XX
DT 12-SEP-1996 (first entry)
XX
DE Alphavirus-based eukaryotic layered vector pVGELVIS.
XX
KW Alphavirus; Sindbis virus; vector; gene therapy; vaccine;
KW pVGELVIS; ss.
XX
OS Chimeric Sindbis virus;
OS Chimeric synthetic.
XX

FH	Key	Location/Qualifiers
FT	promoter	1..60
FT		/*tag= a
FT		/note= "wild-type Sindbis virus sequence
FT		capable of initiating transcription"
FT	CDS	60..5750
FT		/*tag= b
FT		/product= non-structural proteins NSP 1-3
FT	CDS	5928..7579
FT		/*tag= c
FT		/product= non-structural protein NSP 4
FT	misc_feature	7579..7602
FT		/*tag= d
FT		/function= minimal junction region core
FT	misc_feature	11647..11703
FT		/*tag= e
FT		/function= Sindbis polymerase recognition
FT		sequence

XX
PN WO9617072-A2.
XX
PD 06-JUN-1996.
XX
PF 30-NOV-1995; 95WO-US15490.
XX
PR 15-MAR-1995; 95US-0405827.
PR 30-NOV-1994; 94US-0348472.
PR 18-JAN-1995; 95US-0376184.
XX
PA (CHIR) CHIRON VIAGENE INC.
XX
PI Belli BA, Chang SMW, Driver DA, Dubensky TW, Ibanez CE;
PI Jolly DJ, Polo JM;
XX
DR WPI; 1996-277785/28.
XX
PT New recombinant alpha-virus vectors - used to develop prods and
PT methods for use in gene therapy and in the prodn. of vaccines
XX
PS Example 2; Page 186-194; 256pp; English.
XX
CC A eukaryotic layered vector initiation system, pVGELVIS (AAT30787),
CC is based on the Sindbis alphavirus. It is obtd. by inserting
CC the Sindbis clone pVGSP6GENrep into vector pCDNA3. The construct
CC can be used in the prodn. of Sindbis virus-based vectors used
CC to develop products and methods for use in gene therapy and in
CC vaccine prodn. pVGELVIS is capable of initiating alphavirus

```
CC   infection; inactivation of the viral junction region (see also
CC   AAT30810) will prevent viral transcription of the subgenomic
CC   fragment.  The efficiency of pVGELVIS in generating wild-type
CC   Sindbis virus after transfection of BHK cells was approx. 1,000
CC   pfu/mg plasmid DNA.
XX
SQ   Sequence 16656 BP; 4469 A; 4352 C; 4179 G; 3656 T; 0 other;

  Query Match            36.6%;  Score 2370.8;  DB 17;  Length 16656;
  Best Local Similarity  64.6%;  Pred. No. 0;
  Matches 3694;  Conservative    0;  Mismatches 1982;  Indels   44;  Gaps    9;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
              ||||||     |  |  || | ||| || || ||  |||| |||||||  || ||| | |
Db       5746 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5805

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
              || |||| ||.|| |  ||      | ||   |||||| |  ||||| ||     ||||  |
Db       5806 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTCCTGG 5865

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
              | |   | | ||||||   |||||       | ||||  ||    |  |    |     |
Db       5866 AAAGAATTCATGCCCCGGTGCTCGACACGTCGAAAGAGGAACAACTCAAACTCAGGTACC 5925

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
              |  |     || | |||||  || | ||| ||||| || || ||||||||||||||  ||
Db       5926 AGATGATGCCCACCGAAGCCAACAAAAGTAGGTACCAGTCTCGTAAAGTAGAAAATCAGA 5985

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
              |||| || ||  | ||||||||  | || ||| |  |    |||   ||  |   ||
Db       5986 AAGCCATAACCACTGAGCGACTACTGTCAGGACTACGACTGTATAACTCTGCCACAG--- 6042

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
              |||     || || || |   |||  |||||   |||| | |||||    |  |||
Db       6043 ATCAGCCAGAATGCTATAAGATCACCTATCCGAAACCATTGTACTCCAGTAGCGTACCGG 6102

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
              |  |   |  | || || |||| ||       || |||     | | || ||||| |
Db       6103 CGAACTACTCCGATCCACAGTTCGCTGTAGCTGTCTGTAACAACTATCTGCATGAGAACT 6162

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
              | || ||||||||   ||||  || || || || |||||||| ||  | || |||||
Db       6163 ATCCGACAGTAGCATCTTATCAGATTACTGACGAGTACGATGCTTACTTGGATATGGTAG 6222

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
              ||||  ||     ||| || ||||| || || || || || ||||| || ||||| ||||
Db       6223 ACGGGACAGTCGCCTGCCTGGATACTGCAACCTTCTGCCCCGCTAAGCTTAGAAGTTACC 6282

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
              | || || |||  |||  |  ||  |  ||  |    || || |||||  | || ||||
Db       6283 CGAAAAAACATGAGTATAGACCCCCGAATATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6342

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
              | ||  |||||||||||  |   ||| || |||||||| |||||||||| || || ||||
Db       6343 ACACGCTACAAAATGTGCTCATTGCCGCAACTAAAAGAAATTGCAACGTCACGCAGATGC 6402

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
              | ||| | ||     | || || ||| ||||| ||||| || || ||    ||||| ||||
Db       6403 GTGAACTGCCAACACTGGACTCAGCGACATTCAATGTCGAATGCTTTCGAAAATATGCAT 6462

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
              | ||||| ||||| ||||     ||         || ||| || | || || |||   |
Db       6463 GTAATGACGAGTATTGGGAGGAGTTCGCTCGGAAGCCAATTAGGATTACCACTGAGTTTG 6522

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
              | ||   ||||||  | |  |||||||| || || || ||  | ||||| || ||
Db       6523 TCACCGCATATGTAGCTAGACTGAAAGGCCCTAAGGCCGCCACACTATTTGCAAAGACGT 6582

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
              |||||.|    || ||||| ||  | || ||||||  ||||||||||||  | || ||||
Db       6583 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6642

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539

Db       6643 ACGTGAAAGTTACACCAGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6702

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
              | || ||||| ||||| || || |||||| | |||||||| || ||||||||||| ||||
Db       6703 AAGCCGCAGAACCCCTGGCGACTGCTTACTTATGCGGGATTCACCGGGAATTAGTGCGTA 6762

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
              | || |  || ||  | || ||||| || || || || || |||||||| ||||| ||||
Db       6763 GGCTTACGGCCGTCTTGCTTCCAAACATTCACACGCTTTTTGACATGTCGGCGGAGGATT 6822

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
              ||||||| || || || ||||| ||| | || |||||||| ||| |||| ||||| ||||
Db       6823 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGCGACCCGGTACTGGAGACGGATATCG 6882

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
              | || || || |||||| ||||||||||||||  ||  |  | |  ||||||||  | ||||
Db       6883 CATCATTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGTCTGATGATCTTGGAGG 6942

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
              || | ||||| || ||||| ||  | || ||||| |||   || || ||  | ||  |||
Db       6943 ACCTGGGTGTGGATCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 7002

Qy       1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
              |    || |||||||| || ||  | |||||||| || || ||||||||||||||  ||||
Db       7003 CCACCCATCTACCTACGGGTACTCGTTTTAAATTCGGGGCGATGATGAAATCCGGAATGT 7062

Qy       1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
              || | || || |||||||||| | | ||| || | || |||||||||||  |
Db       7063 TCCTCACACTTTTTGTCAACACAGTTTTGAATGTCGTTATCGCCAGCAGAGTACTAGAAG 7122

Qy       1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
              | ||| | |  |||||     || || || |  || |||||||| |||||  | |||||
Db       7123 AGCGGCTTAAAACGTCCAGATGTGCAGCGTTCATTGGCGACGACAACATCATACATGGAG 7182

Qy       2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
              | || || ||||      ||||| ||||| |||||||| ||||| |||||||| || || |
Db       7183 TAGTATCTGACAAAGAAATGGCTGAGAGGTGCGCCACCTGGCTCAACATGGAGGTTAAGA 7242

Qy       2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139
              | || || ||||| || |||  | | |||| |||||||| || |||||||||| ||    |
Db       7243 TCATCGACGCAGTCATCGGTGAGAGACCACCTTACTTCTGCGGCGGATTTATCTTGCAAG 7302

Qy       2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199
              |   | | ||   |||||| ||| | || || || || || |||||||| |||||| | |
Db       7303 ATTCGGTTACTTCCACAGCGTGCCGCGTGGCGGATCCCCTGAAAAGGCTGTTTAAGTTGG 7362

Qy       2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
              | |||||  | |||| ||| ||     ||||||   ||| |  | || || |||| ||||
Db       7363 GTAAACCGCTCCCAGCCGACGACGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 7422

Qy       2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319
              || ||| |   |||   ||| | || || || | |  ||||   ||||||||||   ||
Db       7423 AAACAAAGGCGTGGTTTAGAGTAGGTATAACAGGCACTTTAGCAGTGGCCGTGACGACCC 7482

Qy       2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379
              | || ||| |                | || |  | |  ||  || || ||| | |||
Db       7483 GGTATGAGGTAGACAATATTACACCTGTCCTACTGGCATTGAGAACTTTTGCCCAGAGCA 7542

Qy       2380 TTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGG 2439
                |     ||| |  ||| ||||||     ||      ||||||||  | |||||||||
Db       7543 AAAGAGCATTCCAAGCCATCAGAGGGGAAATAAAGCATCTCTACGGTGGTCCTAAATAGT 7602

Qy       2440 TGACGTAGTAGACACGCACCTAC------------CCACCGGCAGAATGTTTCCATACC 2486
              | |||||  |        | ||                |||||  ||  |      || |
Db       7603 CAGCATAGTACATTTCATCTGACTAATACTACAACACCACCACCATGAATAGAGGATTCT 7662

Qy       2487 CTCAGCTGAACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACC 2546
              | |  || | |  |  |     | |||  |  |  |  |  |  |  |  |  |  |
Db       7663 TTAACATGCTCGGCCGCCGCCCCTTCCCGGCCCCCACTGCCATGTGGAGGCCGCGGAGAA 7722

Qy       2547 CTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTA 2606
                    |    | || |||    ||   | |||||| ||||||||  |  | || |
```


Db 7723 GGAGGCAGGCGGCCCCGATGCCTGCCCGCAACGGGCTGGCTTCTCAAATCCAGCAACTGA 7782
Qy 2607 GGAGGTCGATAGTCAACTTGACTTTCAAACAACGATCACCTAATCCGCCGCCAGGTCCA- 2665
Db 7783 CCACAGCCGTCAGTGCCCTAGTCATTGGACAGGCAACTAGACCTCAACCCCCACGTCCAC 7842
Qy 2666 -----CCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGA 2720
Db 7843 GCCCGCCACCGCGCCAGAAGAAGCAGGCGCCCAAGCAACCACCGAAGCCGAAGAAACCAA 7902
Qy 2721 AGAAGCAGCAAGCCAAGAGGACGAAAC---GCAAGCCTAAACCAGGGAAACGACAACGTA 2777
Db 7903 AAACGCAGGAGAAGAAGAAGAAGCAACCTGCAAAACCCAAACCCGGAAAGAGACAGCGCA 7962
Qy 2778 TGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATCATG---CTGAACGGCCAAGTGA 2834
Db 7963 TGGCACTTAAGTTGGAGGCCGACAGATTGTTCGACGTCAAGAACGAGGACGGAGATGTCA 8022
Qy 2835 ATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACTCCACGTTGAAGGAAAAA 2894
Db 8023 TCGGGCACGCACTGGCCATGGAAGGAAAGGTAATGAAACCTCTGCACGTGAAAGGAACCA 8082
Qy 2895 TTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAGCATGTACGACTTGGAGT 2954
Db 8083 TCGACCACCCTGTGCTATCAAAGCTCAAATTTACCAAGTCGTCAGCATACGACATGGAGT 8142
Qy 2955 ACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTACACCAGCGACAAACCAC 3014
Db 8143 TCGCACAGTTGCCAGTCAACATGAGAAGTGAGGCATTCACCTACACCAGTGAACACCCCG 8202
Qy 3015 CGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAATGGGAGATTTACCGTAC 3074
Db 8203 AAGGATTCTATAACTGGCACCACGGAGCGGTGCAGTATAGTGGAGGTAGATTTACCATCC 8262
Qy 3075 CGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGACAACAGAGGCAGAG 3134
Db 8263 CTCGCGGAGTAGGAGGCAGAGGAGACAGCGGTCGTCCGATCATGGATAACTCCGGTCGGG 8322
Qy 3135 TTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGCGCTTTCAGTGGTCA 3194
Db 8323 TTGTCGCGATAGTCCTCGGTGGCGCTGATGAAGGAACACGAACTGCCCTTTCGGTCGTCA 8382
Qy 3195 CTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTCTGAACCGTGGT--- 3251
Db 8383 CCTGGAATAGTAAAGGGAAGACAATTAAGACAACCCCGGAAGGGACAGAAGAGTGGTCCG 8442
Qy 3252 ------CACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGACAAAC 3305
Db 8443 CAGCACCACTGGTCACGGCAATGTGTTTGCTCGGAAATGTGAGCTTCCCATGCGACCGCC 8502
Qy 3306 CACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAACGTCG 3365
Db 8503 CGCCCACATGCTATACCCGCGAACCTTCCAGAGCCCTCGACATCCTTGAAGAGAACGTGA 8562
Qy 3366 ACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAATGTCCA---TCACGCCGGC 3422
Db 8563 ACCATGAGGCCTACGATACCCTGCTCAATGCCATATTGCGGTGCGGATCGTCTGGCAGAA 8622
Qy 3423 CCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCCCGT 3482
Db 8623 GCAAAAGAAGCGTCGTTGACGACTTTACCCTGACCAGCCCCTACTTGGGCACATGCTCGT 8682
Qy 3483 ATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACGAAT 3542
Db 8683 ACTGCCACCATACTGAACCGTGCTTCAGCCCTGTTAAGATCGAGCAGGTCTGGGACGAAG 8742
Qy 3543 CTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAGGCA 3602
Db 8743 CGGACGATAACACCATACGCATACAGACTTCCGCCCAGTTTGGATACGACCAAAGCGGAG 8802
Qy 3603 CTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAAGGAAG 3662
Db 8803 CAGCAAGCGCAAACAAGTACCGCTACATGTCGCTTAAGCAGGATCACACCGTTAAAGAAG 8862
Qy 3663 ACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCACAAAG 3722
Db 8863 GCACCATGGATGACATCAAGATTAGCACCTCAGGACCGTGTAGAAGGCTTAGCTACAAAG 8922
Qy 3723 GGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCACGAGCG 3782
Db 8923 GATACTTTCTCCTCGCAAAATGCCCTCCAGGGGACAGCGTAACGGTTAGCATAGTGAGTA 8982
Qy 3783 GAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTAGAG 3842
Db 8983 GCAACTCAGCAACGTCATGTACACTGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGG 9042
Qy 3843 AGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATCACT 3902
Db 9043 AAAAATATGATCTACCTCCCGTTCACGGTAAAAGAATTCCTTGCACAGTGTACGACCGTC 9102
Qy 3903 TGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAAGT 3962
Db 9103 TGAAA---ACAACTGCAGGCTACATCACTATGCACAGGCCGGGACCGCACGCTTATACAT 9159
Qy 3963 CCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGTCA 4022
Db 9160 CCTACCTGGAAGAATCATCAGGGAAAGTTTACGCAAAGCCGCCATCTGGGAAGAACATTA 9219
Qy 4023 CCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGATGA 4082
Db 9220 CGTATGAGTGCAAGTGCGGCGACTACAAGACCGGAACCGTTTCGACCCGCACCGAAATCA 9279
Qy 4083 ACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGTCT 4142
Db 9280 CTGGTTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAACGAAGTGGGTCT 9339
Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9340 TCAACTCACCGGACTTGATCAGACATGACGACCACACGGCCCAAGGGAAATTGCATTTGC 9399
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9400 CTTTCAAGTTGATCCCGGGTGCCTGCATGGTCCCTGTTGCCCACGCGCCGAATGTAATAC 9459
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9460 ATGGCTTTAAACACATCAGCCTCCAATTAGATACAGACCACTTGACATTGCTCACCACCA 9519
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9520 GGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCGTCGGAAAGACGGTCAGAA 9579
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9580 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGAAATCATGAGCCAGTGAGGG 9639
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db 9640 TCTATGCCCAAGAGTCAGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9699
Qy 4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db 9700 ATTACTACCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTACCGTGGCGA 9759
Qy 4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db 9760 TGATGATTGGCGTAACTGTTGCAGTGTTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9819
Qy 4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db 9820 CGCCATACGCCCTGGCCCCAAACGCCGTAATCCCAACTTCGCTGGCACTCTTGTGCTGCG 9879
Qy 4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db 9880 TTAGGTCGGCCAATGCTGAAACGTTCACCGAGACCATGAGTTACTTGTGGTCGAACAGTC 9939

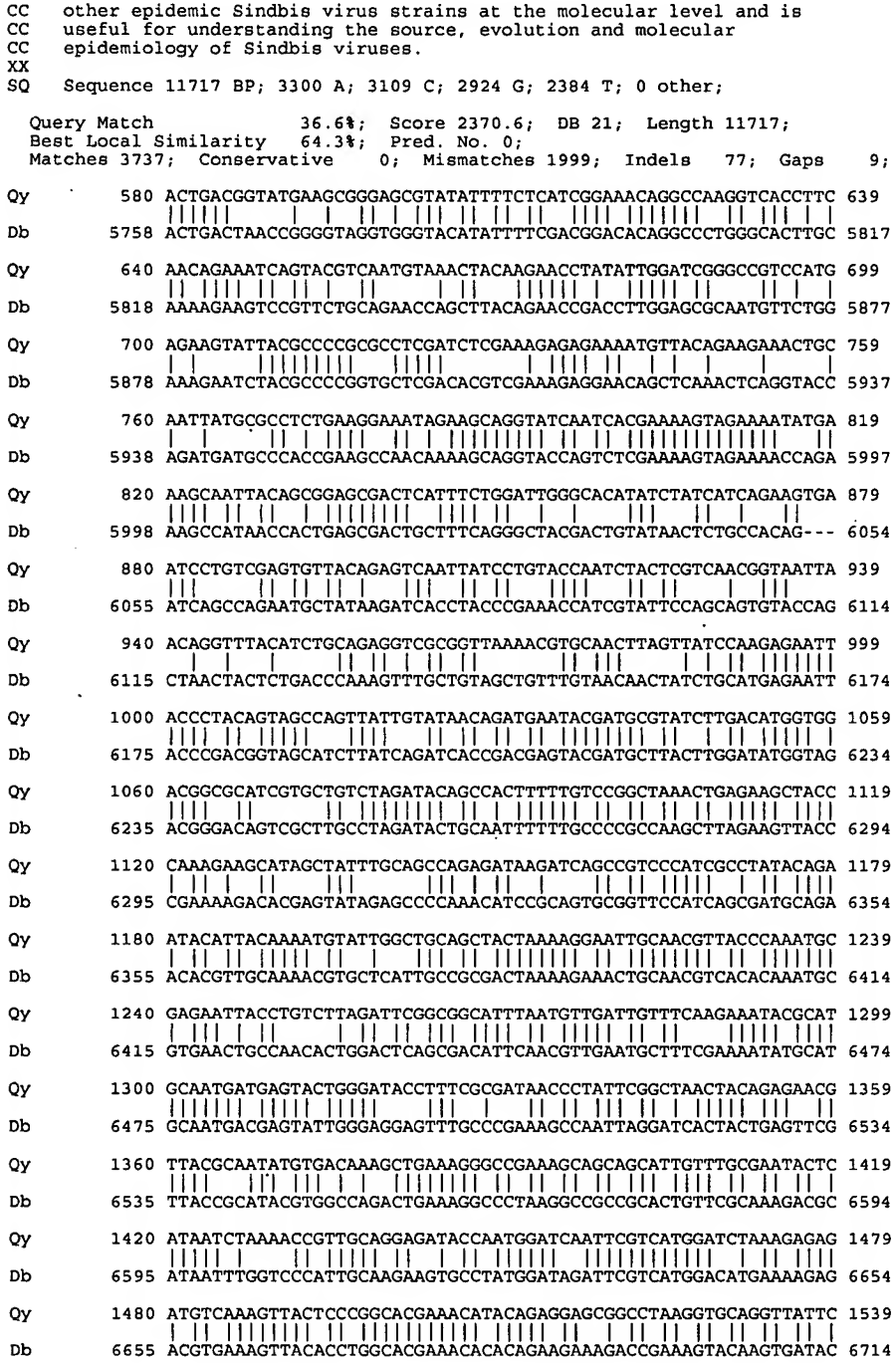
```
Qy       4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
             | |||||  ||||||  ||||||||||| ||| |||| ||| |   |  ||||  | ||||
Db       9940 AGCCGTTCTTCTGGGTCCAGTTGTGCATACCTTTGGCCGCGTTCATCGTTCTAATGCGCT 9999

Qy       4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
              ||  || |||||| |||||||||| ||||||| |||| || ||||| ||||||||||
Db      10000 ACTGCTCCTGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10059

Qy       4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
              ||| ||||||||||||||||| ||||||| ||   ||| ||||||||||  | || |
Db      10060 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10119

Qy       4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
              || | ||||| || || || || ||  |||||||||| ||| |.|| |||||  |    |
Db      10120 AAAGGGCAGGGTATGCCCCGCTCAATTTGGAGATCACTGTCATGTCCTCGGAGGTTTTGC 10179

Qy       4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
              |||| || ||| | |||||| | |||||||||||||  ||| ||  | || ||  ||| ||
Db      10180 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCAAAAA 10239

Qy       5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
              | |||||||||||  ||| | || ||  || |  || |  | || || || || |||   |
Db      10240 TCAAATGCTGCGGCTCCTTGGAATGTCAGCCGGCCGCTCATGCAGACTATACCTGCAAGG 10299

Qy       5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
              | || || || || ||||| || |||||||||||| || ||||| || || ||||||||||
Db      10300 TCTTCGGAGGGGTCTACCCCTTTATGTGGGGAGGAGCGCAATGTTTTTGCGACAGTGAGA 10359

Qy       5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
              |||  ||  |||||||||||||||||| ||  |  |||| ||| |    || |||||
Db      10360 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGCAGATTGCGCGTCTGACCACGCGC 10419

Qy       5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
               ||  | || || ||||| || ||  ||||||| || |||||||||| ||||| ||||
Db      10420 AGGCGATTAAGGTGCACACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10479

Qy       5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
              | |||      ||| |||  ||  || || || ||||| |||||  | ||    ||| |||
Db      10480 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACTTGA 10539

Qy       5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
              | ||||||||  || || || |||||  | ||| | || || || ||||||||||  |||
Db      10540 AAGTCATAGCTGGACCAATTTCAGCATCGTTTACGCCATTCGATCATAAGGTCGTTATCC 10599

Qy       5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
               || || || |||||||| |||||||| || |||||||| ||||||||||||||||
Db      10600 ATCGCGGCCTGGTGTACAACTATGACTTCCCGGAATATGGAGCGATGAAACCAGGAGCGT 10659

Qy       5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
              |||| || ||||||||  ||||  |    |   ||  |   | ||| |||| |||||
Db      10660 TCGGAGACATTCAAGCTACCTCCTTGACTAGCAAGGATCTCATCGCCAGCACAGACATTA 10719

Qy       5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
              |||| || |||||||| | ||||||| | || ||||| ||||| || || | |||||| |
Db      10720 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCAGGAT 10779

Qy       5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
              ||| |||||||| |||||||||| || || ||||| ||||| ||||| || || ||||
Db      10780 TTGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCACCTTTCGGGTGTA 10839

Qy       5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
              | |||| |||  | || || ||||||     ||||| | ||||| |||| || || | |
Db      10840 AGATTGCAGTAAATCCGCTCCGAGCGGTGGACTGTTCATACGGGAACATTCCCATTTCTA 10899

Qy       5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
              |||||||||  | || || ||| | ||  |||||||  |||||     | | |  ||| |
Db      10900 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10959

Qy       5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
               ||   |||    || |||| |||||| |||||||| || ||       ||  | ||||
Db      10960 AATGTGAAGTCAGTGAGTGCACTTATTCAGCAGACTTCGGCGGGATGGCCACCCTGCAGT 11019

Qy       5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
```

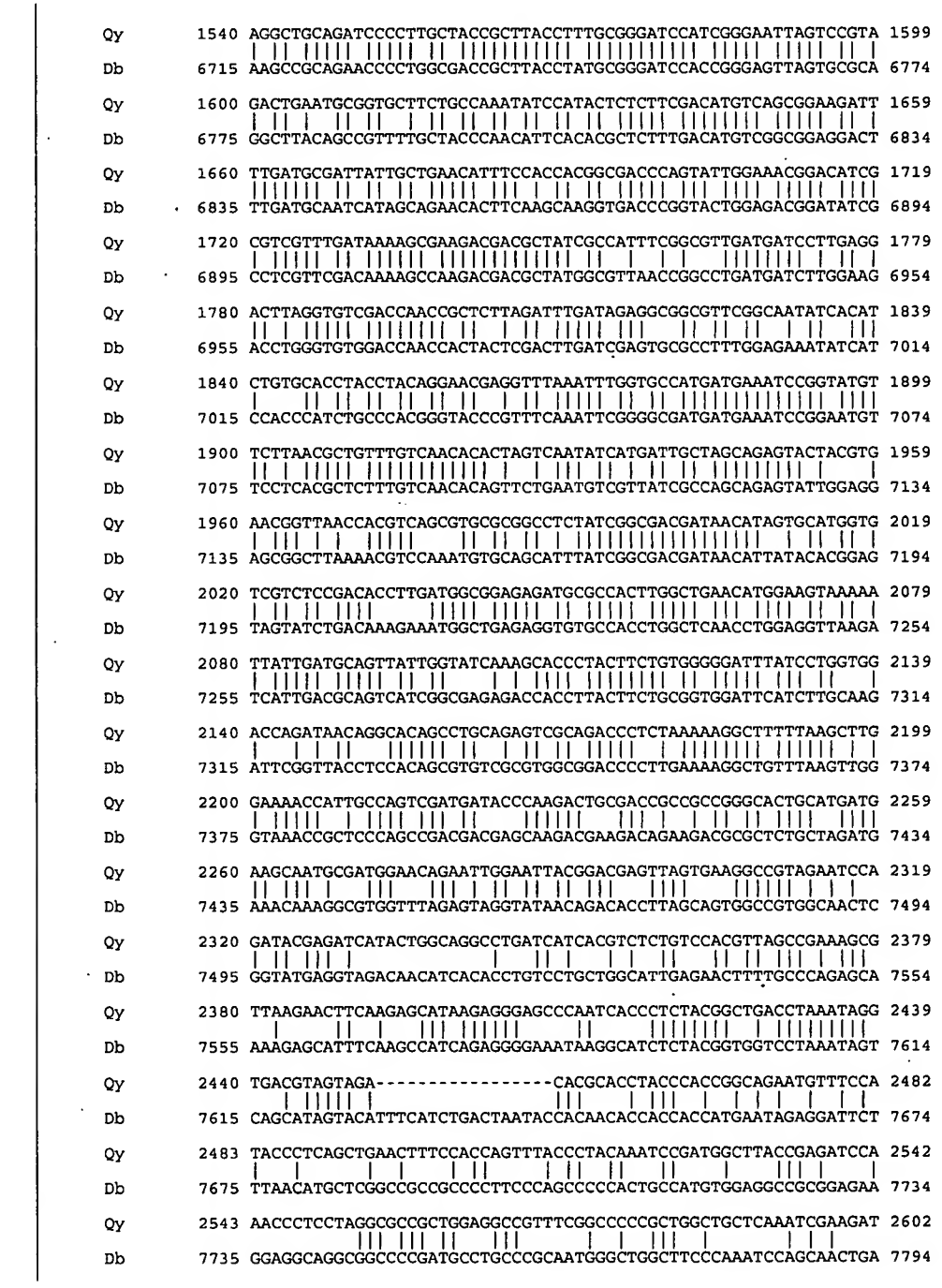
```
             |  | | ||| | || || || || || || || || || || |  |||||   | |
Db      11020 ATGTATCCGACCGCGAAGGTCAATGCCCCGTACATTCGCATTCGAGCACAGCAACTCTCC 11079

Qy       5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
              | ||  |||| ||||||    |    |||      | ||| |||| ||||||||  |||
Db      11080 AAGAGTCGACAGTACATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTTAGCACCGCGA 11139

Qy       5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
              | ||||| || || ||||| || ||||| || || |||||| | || ||||||||  ||||
Db      11140 GTCCACAGGCGAACTTTATCGTATCGCTGTGTGGGAAGAAGACAACATGCAATGCAGAAT 11199

Qy       6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
              |||||||||| || ||||| ||  |  |    || || |||   |||||||||| || |
Db      11200 GTAAACCACCAGCTGACCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTTCAAG 11259

Qy       6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
              | ||  | || |||||||| |||| |||||| |||| || || || || || || || |
Db      11260 CCGCCATCTCAAAAACATCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11319

Qy       6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
              |  |  || ||||||||||  | || |  ||||||   ||||| |  | ||||||  ||||
Db      11320 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11379

Qy       6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
              ||| |          || ||| |  | | | ||||||||||||||||||||||  |  ||
Db      11380 GACCGCTACGCCCCAATGATCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11439

Qy       6242 GCATAATGCCACGCGCCGCT 6261
              |||||||||  |  || | |
Db      11440 GCATAATGCATCAGGCTGGT 11459
```

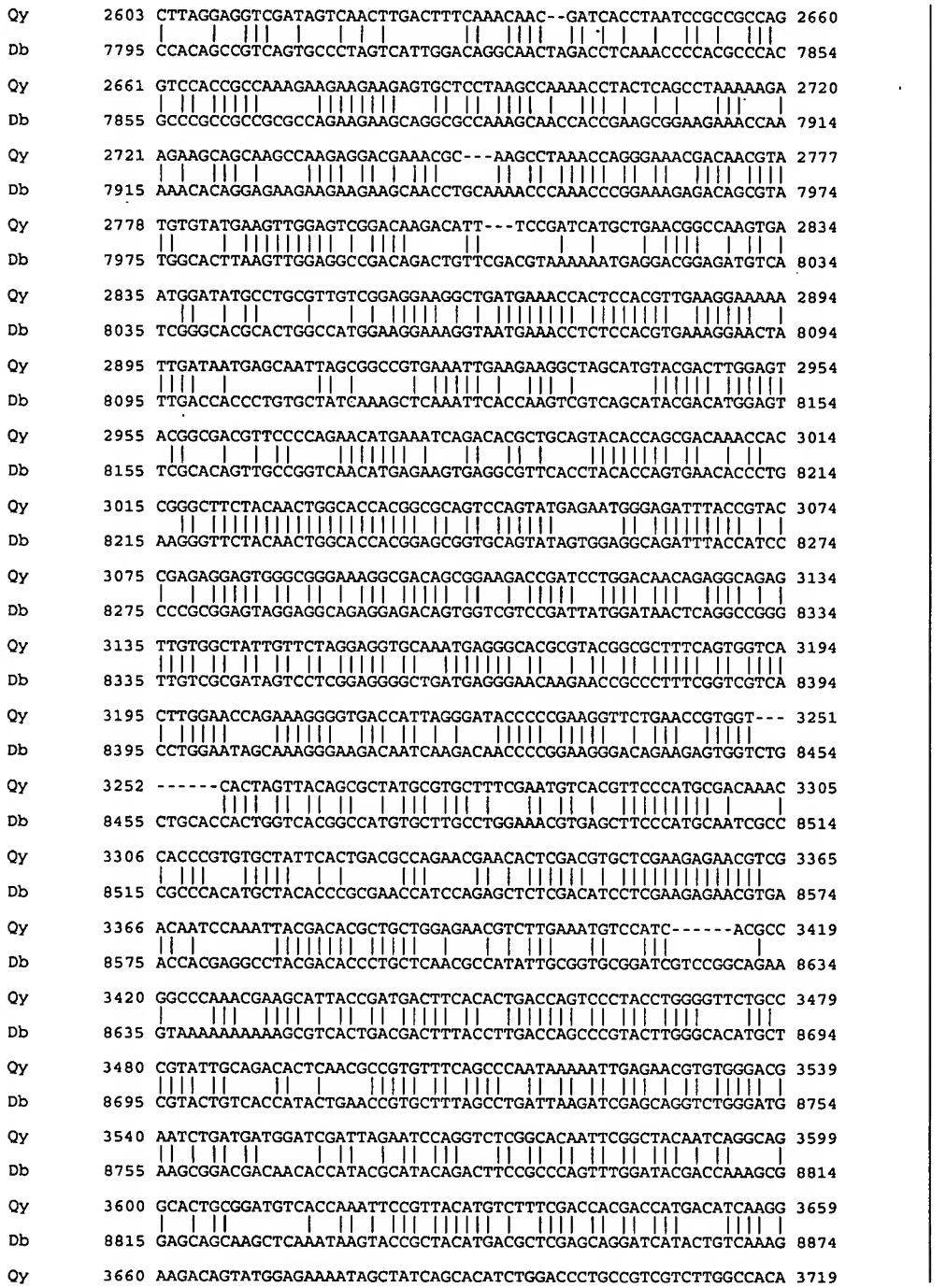
RESULT 11
AAA70577
ID AAA70577 standard; cDNA; 11717 BP.
XX
AC AAA70577;
XX
DT 06-DEC-2000 (first entry)
XX
DE Sindbis-like virus strain YN87448 complete genome.
XX
KW Genome; Sindbis-like virus strain YN87448; primer; RT-PCR; vaccine;
KW epidemic; Sindbis encephalitis; evolution; epidemiology; ds.
XX
OS Sindbis-like virus strain YN87448.
XX
PN CN1252445-A.
XX
PD 10-MAY-2000.
XX
PF 27-OCT-1998; 98CN-0120694.
XX
PR 27-OCT-1998; 98CN-0120694.
XX
PA (VIRO-) INST VIROLOGY CHINESE ACAD PREVENTIVE ME.
XX
PI Liang G, Zhou G, Li L;
XX
DR WPI; 2000-443226/39.
XX
PT Whole genome sequence of YN87448 virus strain and its cloning method -
XX
PS Claim 1; Page 2-9; 24pp; Chinese.
XX
CC This sequence represents the complete genome of the Sindbis-like virus
CC strain YN87448. The genome was cloned as 15 fragments using the PCR
CC primers AAA70578-A70603 into the plasmid pGEM-T. The invention relates
CC to the isolation and method of cloning the complete genome for the
CC Sindbis-like virus strain YN87448 by a RT-PCR process. The YN87448
CC strain virus appears to be the optimal candidate for a vaccine to
CC prevent epidemics of Sindbis encephalitis. The sequence of this
CC strain's genome shows the difference between this viral strain and

CC other epidemic Sindbis virus strains at the molecular level and is
CC useful for understanding the source, evolution and molecular
CC epidemiology of Sindbis viruses.
XX
SQ Sequence 11717 BP; 3300 A; 3109 C; 2924 G; 2384 T; 0 other;

Query Match 36.6%; Score 2370.6; DB 21; Length 11717;
Best Local Similarity 64.3%; Pred. No. 0;
Matches 3737; Conservative 0; Mismatches 1999; Indels 77; Gaps 9;






Qy 2603 CTTAGGAGGTCGATAGTCAACTTGACTTTCAAACAAC--GATCACCTAATCCGCCGCCAG 2660
Db 7795 CCACAGCCGTCAGTGCCCTAGTCATTGGACAGGCAACTAGACCTCAAACCCCACGCCCAC 7854
Qy 2661 GTCCACCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGA 2720
Db 7855 GCCCGCCGCCGCGCCAGAAGAAGCAGGCGCCAAAGCAACCACCGAAGCGGAAGAAACCAA 7914
Qy 2721 AGAAGCAGCAAGCCAAGAGGACGAAACGC---AAGCCTAAACCAGGGAAACGACAACGTA 2777
Db 7915 AAACACAGGAGAAGAAGAAGAAGCAACCTGCAAAACCCAAACCCGGAAAGAGACAGCGTA 7974
Qy 2778 TGTGTATGAAGTTGGAGTCGGACAAGACATT---TCCGATCATGCTGAACGGCCAAGTGA 2834
Db 7975 TGGCACTTAAGTTGGAGGCCGACAGACTGTTCGACGTAAAAAATGAGGACGGAGATGTCA 8034
Qy 2835 ATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACTCCACGTTGAAGGAAAAA 2894
Db 8035 TCGGGCACGCACTGGCCATGGAAGGAAAGGTAATGAAACCTCTCCACGTGAAAGGAACTA 8094
Qy 2895 TTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAGCATGTACGACTTGGAGT 2954
Db 8095 TTGACCACCCTGTGCTATCAAAGCTCAAATTCACCAAGTCGTCAGCATACGACATGGAGT 8154
Qy 2955 ACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTACACCAGCGACAAACCAC 3014
Db 8155 TCGCACAGTTGCCGGTCAACATGAGAAGTGAGGCGTTCACCTACACCAGTGAACACCCTG 8214
Qy 3015 CGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAATGGGAGATTTACCGTAC 3074
Db 8215 AAGGGTTCTACAACTGGCACCACGGAGCGGTGCAGTATAGTGGAGGCAGATTTACCATCC 8274
Qy 3075 CGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGACAACAGAGGCAGAG 3134
Db 8275 CCCGCGGAGTAGGAGGCAGAGGAGACAGTGGTCGTCCGATTATGGATAACTCAGGCCGGG 8334
Qy 3135 TTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGCGCTTTCAGTGGTCA 3194
Db 8335 TTGTCGCGATAGTCCTCGGAGGGGCTGATGAGGGAACAAGAACCGCCCTTTCGGTCGTCA 8394
Qy 3195 CTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTCTGAACCGTGGT--- 3251
Db 8395 CCTGGAATAGCAAAGGGAAGACAATCAAGACAACCCCGGAAGGGACAGAAGAGTGGTCTG 8454
Qy 3252 ------CACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGACAAAC 3305
Db 8455 CTGCACCACTGGTCACGGCCATGTGCTTGCCTGGAAACGTGAGCTTCCCATGCAATCGCC 8514
Qy 3306 CACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAACGTCG 3365
Db 8515 CGCCCACATGCTACACCCGCGAACCATCCAGAGCTCTCGACATCCTCGAAGAGAACGTGA 8574
Qy 3366 ACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAATGTCCATC------ACGCC 3419
Db 8575 ACCACGAGGCCTACGACACCCTGCTCAACGCCATATTGCGGTGCGGATCGTCCGGCAGAA 8634
Qy 3420 GGCCCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCC 3479
Db 8635 GTAAAAAAAAAAGCGTCACTGACGACTTTACCTTGACCAGCCCGTACTTGGGCACATGCT 8694
Qy 3480 CGTATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACG 3539
Db 8695 CGTACTGTCACCATACTGAACCGTGCTTTAGCCTGATTAAGATCGAGCAGGTCTGGGATG 8754
Qy 3540 AATCTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAG 3599
Db 8755 AAGCGGACGACAACACCATACGCATACAGACTTCCGCCCAGTTTGGATACGACCAAAGCG 8814
Qy 3600 GCACTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAAGG 3659
Db 8815 GAGCAGCAAGCTCAAATAAGTACCGCTACATGACGCTCGAGCAGGATCATACTGTCAAAG 8874
Qy 3660 AAGACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCACA 3719
Db 8875 AAGGCACCATGGATGACATCAAGATCAGCACCTCAGGACCGTGTGGAAGGCTTAGCTACA 8934
Qy 3720 AAGGGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCACGA 3779
Db 8935 AAGGATACTTTCTCCTCGCGAAGTGTCCTCCAGGGGACAGCGTAACGGTTAGCATAGCGA 8994
Qy 3780 GCGGAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTA 3839
Db 8995 GTAGCAACTCAGCAACGTCATGCACAATGGCCCGCAAGATAAAACCAAAATTCGTGGGAC 9054
Qy 3840 GAGAGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATC 3899
Db 9055 GGGAAAAATATGACCTACCTCCCGTTCACGGTAAGAAGATTCCTTGCACAGTGTACGACC 9114
Qy 3900 ACTTGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATA 3959
Db 9115 GTCTGAAAGAAACAACCGCCGGCTACATCACTATGCACAGGCCGGGACCGCACGCCTATA 9174
Qy 3960 AGTCCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACG 4019
Db 9175 CATCCTATCTGAAGGAATCATCAGGGAAAGTTTACGCGAAGCCACCATCCGGGAAGAACA 9234
Qy 4020 TCACCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGA 4079
Db 9235 TTACGTACGAGTGCAAGTGCGGCGATTACAAGACCGGAACCGTTACGACCCGTACCGAAA 9294
Qy 4080 TGAACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGG 4139
Db 9295 TCACGGGCTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAACGAAGTGGG 9354
Qy 4140 TCTTCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACA 4199
Db 9355 TCTTCAACTCGCCGGACTCTATCAGACACGCCGACCACACGGCCCAAGGGAAATTGCATT 9414
Qy 4200 TTCCATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCA 4259
Db 9415 TGCCTTTCAAGCTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAACGTAG 9474
Qy 4260 CGAAGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAA 4319
Db 9475 TACACGGCTTTAAACACATCAGCCTCCAATTAGACACAGACCATCTGACATTGCTCACCA 9534
Qy 4320 CGAGAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCA 4379
Db 9535 CCAGGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCATCGGAAAGACGGTTA 9594
Qy 4380 GGAATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCA 4439
Db 9595 GAAACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGCAATCACGAACCAGTAA 9654
Qy 4440 GAGTCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCA 4499
Db 9655 GGGTCTATGCCCAAGAGTCTGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTAC 9714
Qy 4500 TCCACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTG 4559
Db 9715 AGCATTACTATCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTGCTGTGG 9774
Qy 4560 CTATCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCC 4619
Db 9775 CGATGATGATTGGCGTAACTGTTGCAGCATTATGTGCCTGTAAAGCGCGCCGTGAGTGCC 9834
Qy 4620 TGACGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCT 4679
Db 9835 TGACGCCATATGCCCTGGCCCCAAATGCCGTGATTCCAACTTCGCTGGCACTTTTGTGCT 9894
Qy 4680 GCATTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACA 4739
Db 9895 GTGTTAGGTCGGCTAATGCTGAAACATTCACCGAGACCATGAGTTACTTATGGTCGAACA 9954
Qy 4740 ACCAACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCC 4799

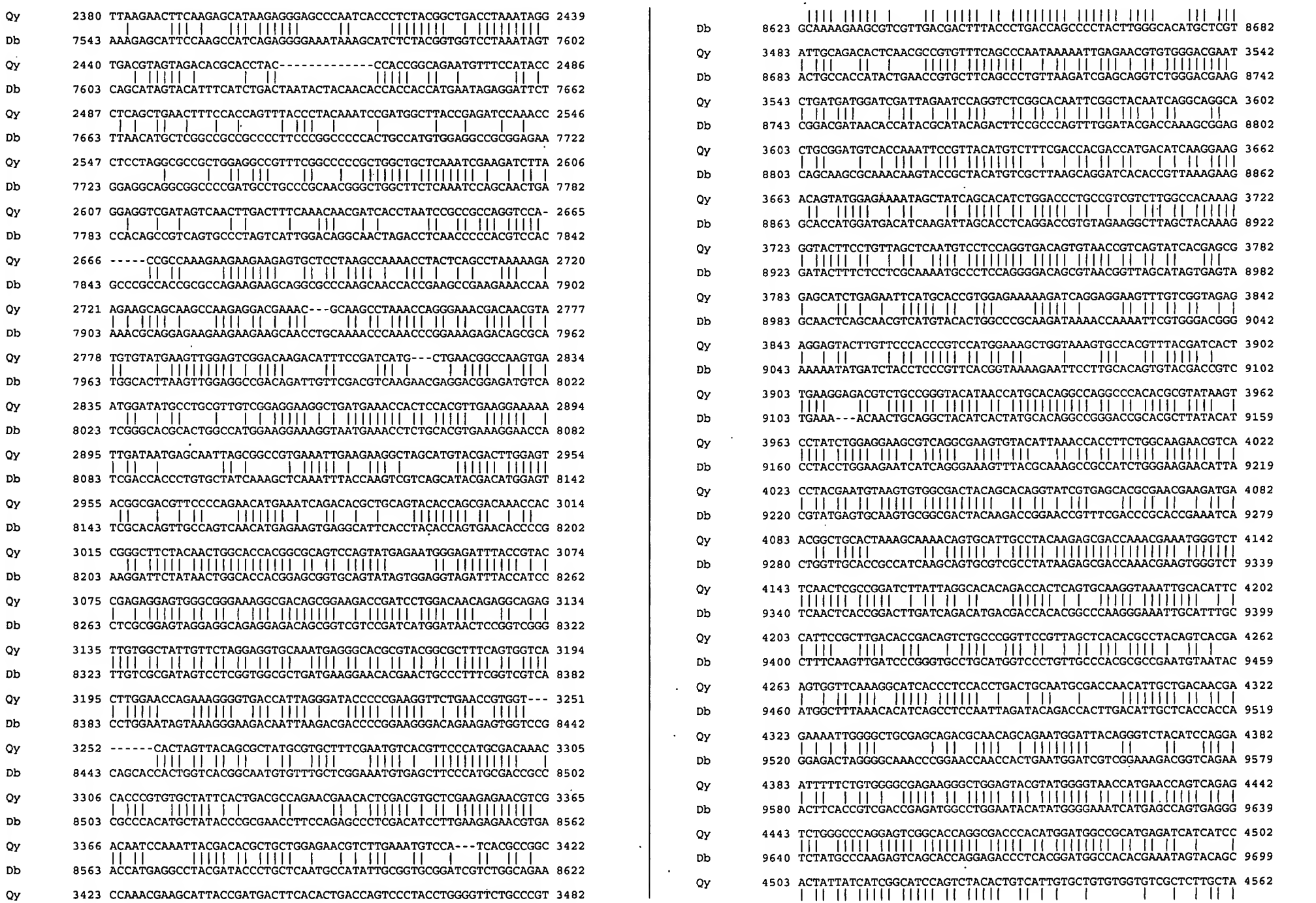
```
Db       9955 GCCAGCCGTTCTTCTGGGTCCAGCTGTGTATACCTCTGGCCGCTGTCATCGTTCTAATGC 10014

Qy       4800 GCTGCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAG 4859

Db      10015 GCTGTTGCTCATGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAG 10074

Qy       4860 ACGCCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGG 4919

Db      10075 ACGCCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCATATAAGGCACTTG 10134

Qy       4920 TCGAACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAA 4979

Db      10135 TTGAAAGGGCAGGGTACGCCCCGCTCAGTTTGGAGATTACTGTCATGTCCTCGTAGGTTT 10194

Qy       4980 CACCTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCAC 5039

Db      10195 TGCCTTCCACCAACCAAGAGTACATCACCTGCAAATTCACCACTGTGGTCCCCTCCCCTA 10254

Qy       5040 AAGTTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCC 5099

Db      10255 AAGTCAGATGCTGCGGCTCCTTGGAATGTCAGCCCGCGGCTCACGCAGATTATACCTGCA 10314

Qy       5100 GCGTTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTG 5159

Db      10315 AGGTCTTTGGAGGGGTGTACCCCTTCATGTGGGGAGGAGCACAATGTTTTTGCGACAGTG 10374

Qy       5160 AGAACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACG 5219

Db      10375 AGAACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGCAGATTGCGCGACTGACCACG 10434

Qy       5220 CAGTCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCA 5279

Db      10435 CGCAGGCGATTAAGGTGCATACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGA 10494

Qy       5280 ACACCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACC 5339

Db      10495 ACACTATCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACC 10554

Qy       5340 TGAAGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCA 5399

Db      10555 TGAAAGTCATAGCTGGACCAATTTCAGCATTGTTTACACCATTCGATCACAAGGTCGTTA 10614

Qy       5400 TCAGAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAG 5459

Db      10615 TCCATCGCGGCCTGGTGTACAACTATGACTTTCCGGAATACGGAGCGATGAAACCAGGAG 10674

Qy       5460 CGTTCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACA 5519

Db      10675 CGTTTGGAGACATTCAAGCTACCTCCTTGACTAGCAAAGACCTCATCGCCAGCACAGATA 10734

Qy       5520 TACGGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAG 5579

Db      10735 TTAGGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCTG 10794

Qy       5580 GGTATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGAT 5639

Db      10795 GATTCGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCCCCTTTTGGGC 10854

Qy       5640 GTAAAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCT 5699

Db      10855 GCAAGATTGCAGTCAATCCGCTTCAGGCGGTGGACTGCTCATACGGGAACATTCCCATCT 10914

Qy       5700 CGATTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAG 5759

Db      10915 CTATTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAG 10974

Qy       5760 TTAGCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTAC 5819

Db      10975 TCAAATGTGATGTCAGTGAGTGCACTTATTCAGCGGACTTCGGCGGGATGGCTACCCTGC 11034

Qy       5820 AGTACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTT 5879

Db      11035 AGTATGTATCCGACCGCGAAGGACAATGCCCTGTACATTCGCATTCGAGCACAGCAACCC 11094

Qy       5880 TGAAGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACAT 5939

Db      11095 TCCAAGAGTCGACAGTTCATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTCAGCACCG 11154

Qy       5940 CGAGCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTG 5999

Db      11155 CGAGCCCACAGGCAAACTTCATTGTATCGCTGTGTGGTAAGAAGACAACATGCAATGCAG 11214

Qy       6000 AATGTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCC 6059

Db      11215 AATGCAAACCACCAGCTGATCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTCC 11274

Qy       6060 AGGCGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCAT 6119

Db      11275 AAGCCGCCATCTCAAAAACTTCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGT 11334

Qy       6120 CCCTCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTA 6179

Db      11335 CGCTATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAA 11394

Qy       6180 GATGACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGT 6238

Db      11395 GATGACCGCTACGCCCCAATGACCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGA 11454

Qy       6239 GGTGCATAATGCCACGCGCCGCT--------------------------------TGAC 6265

Db      11455 TGTGCATAATGCATCAGGCTGGTATATTAGATCCCCGCTTACCGCGGGCAATATAGCAAC 11514

Qy       6266 ACTAAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGT 6318

Db      11515 ACCAAAACTCGACGTATTTCCGAGGAAGCGCAGTGCATAATGCTGCGCAGTGT 11567
```

RESULT 12
AAV60125
ID AAV60125 standard; DNA; 16656 BP.
XX
AC AAV60125;
XX
DT 25-MAR-2003 (updated)
DT 04-DEC-1998 (first entry)
XX
DE Representative eukaryotic layered vector initiation system.
XX
KW Eukaryotic layered vector initiation system; stimulate; immune response;
KW ss.
XX
OS Synthetic.
OS Sindbis virus.
XX
PN US5814482-A.
XX
PD 29-SEP-1998.
XX
PF 30-OCT-1996; 96US-0739158.
XX
PR 14-SEP-1994; 94WO-US10469.
PR 15-MAR-1995; 95US-0404796.
PR 15-SEP-1993; 93US-0122791.
PR 18-FEB-1994; 94US-0198450.
PR 30-NOV-1994; 94US-0348472.
PR 18-JAN-1995; 95US-0376184.
XX
PA (DRIV/) DRIVER D A.
PA (DUBE/) DUBENSKY T W.
PA (JOLL/) JOLLY D J.
PA (POLO/) POLO J M.
XX
PI Driver DA, Dubensky TW, Jolly DJ, Polo JM;
XX
DR WPI; 1998-541753/46.
XX

PT Eukaryotic layered vector initiation system - containing eukaryotic
PT promoter and heterologous antigen coding sequence, useful for
PT stimulating immune response
XX
PS Example 1; Fig 3A-H; 144pp; English.
XX
CC The present sequence is a representative eukaryotic layered vector
CC initiation system, based on Sindbis. The eukaryotic layered vector
CC initiation system comprises a eukaryotic promoter 5' of viral cDNA
CC which initiates, in a susceptible target cell, 5' to 3' synthesis of
CC RNA from the viral cDNA. The RNA comprises a vector construct which
CC autonomously amplifies in the cell and expresses a heterologous nucleic
CC acid sequence which encodes an antigen or modified form that stimulates
CC an immune response within an animal. The system is useful for stimulating
CC an immune response to an antigen by introducing the vector into target
CC cells, preferably by infection in vivo, especially where the immune
CC response is a cell mediated, HLA class I-restricted or an HLA class
CC II-restricted immune response.
CC (Updated on 25-MAR-2003 to correct PR field.)
XX
SQ Sequence 16656 BP; 4467 A; 4349 C; 4183 G; 3657 T; 0 other;

Query Match 36.5%; Score 2369.2; DB 19; Length 16656;
Best Local Similarity 64.6%; Pred. No. 0;
Matches 3693; Conservative 0; Mismatches 1983; Indels 44; Gaps 9;

```
Qy       580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
Db      5746 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5805

Qy       640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
Db      5806 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTCCTGG 5865

Qy       700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
Db      5866 AAAGAATTCATGCCCCGGTGCTCGACACGTCGAAAGAGGAACAACTCAAACTCAGGTACC 5925

Qy       760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
Db      5926 AGATGATGCCCACCGAAGCCAACAAAAGTAGGTACCAGTCTCGTAAAGTAGAAAATCAGA 5985

Qy       820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
Db      5986 AAGCCATAACCACTGAGCGACTACTGTCAGGACTACGACTGTATAACTCTGCCACAG--- 6042

Qy       880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
Db      6043 ATCAGCCAGAATGCTATAAGATCACCTATCCGAAACCATTGTACTCCAGTAGCGTACCGG 6102

Qy       940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
Db      6103 CGAACTACTCCGATCCACAGTTCGCTGTAGCTGTCTGTAACAACTATCTGCATGAGAACT 6162

Qy      1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
Db      6163 ATCCGACAGTAGCATCTTATCAGATTACTGACGAGTACGATGCTTACTTGGATATGGTAG 6222

Qy      1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
Db      6223 ACGGGACAGTCGCCTGCCTGGATACTGCAACCTTCTGCCCCGCTAAGCTTAGAAGTTACC 6282

Qy      1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
Db      6283 CGAAAAAACATGAGTATAGAGCCCCGAATATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6342

Qy      1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
Db      6343 ACACGCTACAAAATGTGCTCATTGCCGCAACTAAAAGAAATTGCAACGTCACGCAGATGC 6402

Qy      1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
Db      6403 GTGAACTGCCAACACTGGACTCAGCGACATTCAATGTCGAATGCTTTCGAAAATATGCAT 6462

Qy      1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
Db      6463 GTAATGACGAGTATTGGGAGGAGTTCGCTCGGAAGCCAATTAGGATTACCACTGAGTTTG 6522

Qy      1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
Db      6523 TCACCGCATATGTAGCTAGACTGAAAGGCCCTAAGGCCGCCACACTATTTGCAAAGACGT 6582

Qy      1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
Db      6583 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6642

Qy      1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
Db      6643 ACGTGAAAGTTACACCAGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6702

Qy      1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
Db      6703 AAGCCGCAGAACCCCTGGCGACTGCTTACTTATGCGGGATTCACCGGGAATTAGTGCGTA 6762

Qy      1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
Db      6763 GGCTTACGGCCGTCTTGCTTCCAAACATTCACACGCTTTTTGACATGTCGGCGGAGGATT 6822

Qy      1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
Db      6823 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGCGACCCGGTACTGGAGACGGATATCG 6882

Qy      1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
Db      6883 CATCATTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGTCTGATGATCTTGGAGG 6942

Qy      1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
Db      6943 ACCTGGGTGTGGATCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 7002

Qy      1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
Db      7003 CCACCCATCTACCTACGGGTACTCGTTTTAAATTCGGGGCGATGATGAAATCCGGAATGT 7062

Qy      1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
Db      7063 TCCTCACACTTTTTGTCAACACAGTTTGAATGTCGTTATCGCCAGCAGAGTACTAGAAG 7122

Qy      1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
Db      7123 AGCGGCTTAAAACGTCCAGATGTGCAGCGTTCATTGGCGACGACAACATCATACATGGAG 7182

Qy      2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
Db      7183 TAGTATCTGACAAAGAAATGGCTGAGAGGTGCGCCACCTGGCTCAACATGGAGGTTAAGA 7242

Qy      2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139
Db      7243 TCATCGACGCAGTCATCGGTGAGAGACCACCTTACTTCTGCGGCGGATTTATCTTGCAAG 7302

Qy      2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199
Db      7303 ATTCGGTTACTTCCACAGCGTGCCGCGTGGCGGATCCCCTGAAAAGGCTGTTTAAGTTGG 7362

Qy      2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
Db      7363 GTAAACCGCTCCCAGCCGACGACGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 7422

Qy      2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319
Db      7423 AAACAAAGGCGTGGTTTAGAGTAGGTATAACAGGCACTTTAGCAGTGGCCGTGACGACCC 7482

Qy      2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379
Db      7483 GGTATGAGGTAGACAATATTACACCTGTCCTACTGGCATTGAGAACTTTTGCCCAGAGCA 7542
```



```
Db      9700 ATTACTACCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTACCGTGGCGA 9759

Qy      4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622

Db      9760 TGATGATTGGCGTAACTGTTGCAGTGTTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9819

Qy      4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682

Db      9820 CGCCATACGCCCTGGCCCCAAACGCCGTAATCCCAACTTCGCTGGCACTCTTGTGCTGCG 9879

Qy      4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742

Db      9880 TTAGGTCGGCCAATGCTGAAACGTTCACCGAGACCATGAGTTACTTGTGGTCGAACAGTC 9939

Qy      4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802

Db      9940 AGCCGTTCTTCTGGGTCCAGTTGTGCATACCTTTGGCCGCGTTCATCGTTCTAATGCGCT 9999

Qy      4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862

Db     10000 ACTGCTCCTGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10059

Qy      4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922

Db     10060 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10119

Qy      4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982

Db     10120 AAAGGGCAGGGTATGCCCCGCTCAATTTGGAGATCACTGTCATGTCCTCGGAGGTTTTGC 10179

Qy      4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042

Db     10180 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCAAAAA 10239

Qy      5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102

Db     10240 TCAAATGCTGCGGCTCCTTGGAATGTCAGCCGGCCGCTCATGCAGACTATACCTGCAAGG 10299

Qy      5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162

Db     10300 TCTTCGGAGGGGTCTACCCCTTTATGTGGGGAGGAGCGCAATGTTTTTGCGACAGTGAGA 10359

Qy      5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222

Db     10360 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGCAGATTGCGCGTCTGACCACGCGC 10419

Qy      5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282

Db     10420 AGGCGATTAAGGTGCACACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10479

Qy      5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342

Db     10480 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACTTGA 10539

Qy      5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402

Db     10540 AAGTCATAGCTGGACCAATTTCAGCATCGTTTACGCCATTCGATCATAAGGTCGTTATCC 10599

Qy      5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462

Db     10600 ATCGCGGCCTGGTGTACAACTATGACTTCCCGGAATATGGAGCGATGAAACCAGGAGCGT 10659

Qy      5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522

Db     10660 TCGGAGACATTCAAGCTACCTCCTTGACTAGCAAGGATCTCATCGCCAGCACAGACATTA 10719

Qy      5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582

Db     10720 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCAGGAT 10779

Qy      5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642

Db     10780 TTGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCACCTTTCGGGTGTA 10839

Qy      5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702

Db     10840 AGATTGCAGTAAATCCGCTCCGAGCGGTGGACTGTTCATACGGGAACATTCCCATTTCTA 10899

Qy      5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762

Db     10900 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10959

Qy      5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822

Db     10960 AATGTGAAGTCAGTGAGTGCACTTATTCAGCAGACTTCGGCGGGATGGCCACCCTGCAGT 11019

Qy      5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882

Db     11020 ATGTATCCGACCGCGAAGGTCAATGCCCCGTACATTCGCATTCGAGCACAGCAACTCTCC 11079

Qy      5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942

Db     11080 AAGAGTCGACAGTACATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTTAGCACCGCGA 11139

Qy      5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002

Db     11140 GTCCACAGGCGAACTTTATCGTATCGCTGTGTGGGAAGAAGACAACATGCAATGCAGAAT 11199

Qy      6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062

Db     11200 GTAAACCACCAGCTGACCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTTCAAG 11259

Qy      6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122

Db     11260 CCGCCATCTCAAAAACATCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11319

Qy      6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182

Db     11320 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11379

Qy      6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241

Db     11380 GACCGCTACGCCCCAATGATCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11439

Qy      6242 GCATAATGCCACGCGCCGCT 6261

Db     11440 GCATAATGCATCAGGCTGGT 11459
```

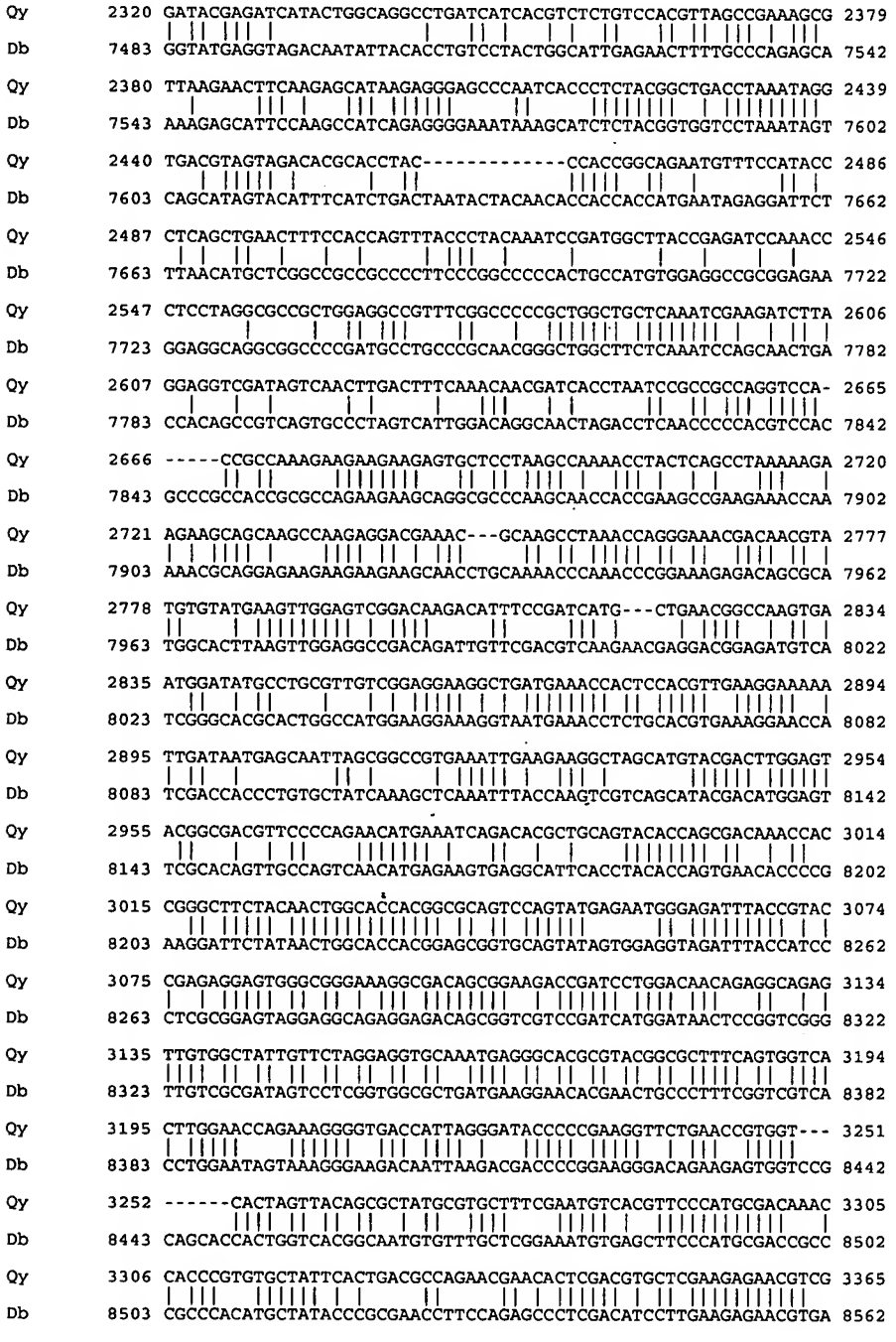
RESULT 13
AAV42364
ID AAV42364 standard; DNA; 16656 BP.
XX
AC AAV42364;
XX
DT 02-OCT-1998 (first entry)
XX
DE Representative eukaryotic layered vector initiation system sequence.
XX
KW Representative eukaryotic layered vector initiation system;
KW DNA alphavirus; structural protein expression; inhibit; pathogen;
KW immune response; stimulate; ss.
XX
OS Sindbis virus.
XX
PN US5789245-A.
XX
PD 04-AUG-1998.
XX
PF 30-OCT-1996; 96US-0741881.
XX
PR 15-MAR-1995; 95US-0404796.
PR 15-SEP-1993; 93US-0122791.
PR 18-FEB-1994; 94US-0198450.
PR 30-NOV-1994; 94US-0348472.
PR 20-JAN-1995; 95US-0376184.

PR 30-OCT-1996; 96US-0741881.
XX
PA (CHIR) CHIRON CORP.
XX
PI Chang SMW, Driver DA, Dubensky TW, Ibanez CE, Jolly DJ;
PI Polo JM;
XX
DR WPI; 1998-446089/38.
XX
PT DNA alpha:virus structural protein expression cassettes - for
PT producing recombinant alpha:virus particles
XX
PS Disclosure; Fig 3A-H; 140pp; English.
XX
CC The present sequence is a representative eukaryotic layered vector
CC initiation system derived from Sindbis virus. The specification describes
CC a DNA alphavirus structural protein expression cassette which comprises
CC an inducible promoter and an alphavirus structural protein gene, where
CC the promoter directs the expression of the alphavirus structural protein
CC gene upon induction of the promoter within a cell, and where prior to
CC induction within the cell, the expression cassette does not express
CC sufficient quantities of structural proteins to be cytotoxic to a BHK
CC cell containing the expression cassette. The products may be used to
CC inhibit pathogens and stimulate an immune response.
XX
SQ Sequence 16656 BP; 4467 A; 4349 C; 4183 G; 3657 T; 0 other;

Query Match 36.5%; Score 2369.2; DB 19; Length 16656;
Best Local Similarity 64.6%; Pred. No. 0;
Matches 3693; Conservative 0; Mismatches 1983; Indels 44; Gaps 9;

```
Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
              ||||||     |  |  || | ||| || || ||  |||| |||||||  || ||| | |
Db       5746 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5805

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
              || |||| || || |  ||      | ||   ||||||  |  |||||  ||    ||||  |
Db       5806 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTCCTGG 5865

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
              | |    | | ||||||   |||||        | |||| ||     |  |      |
Db       5866 AAAGAATTCATGCCCCGGTGCTCGACACGTCGAAAGAGGAACAACTCAAACTCAGGTACC 5925

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
              |  |     || | ||||  || | ||| |||||  || || || ||||||||||||  ||
Db       5926 AGATGATGCCCACCGAAGCCAACAAAAGTAGGTACCAGTCTCGTAAAGTAGAAAATCAGA 5985

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
              |||| || ||  | |||||||  | || ||| |  |    |||   ||  |   ||
Db       5986 AAGCCATAACCACTGAGCGACTACTGTCAGGACTACGACTGTATAACTCTGCCACAG--- 6042

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
              |||     || || |   |||  |||||   |||| | |||||     |  |||
Db       6043 ATCAGCCAGAATGCTATAAGATCACCTATCCGAAACCATTGTACTCCAGTAGCGTACCGG 6102

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
                |  |   |  | || || |||| ||       || |||     | | || |||||  |
Db       6103 CGAACTACTCCGATCCACAGTTCGCTGTAGCTGTCTGTAACAACTATCTGCATGAGAACT 6162

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
              | || ||||||||   ||||   || || || || |||||||| ||  | || |||||  |
Db       6163 ATCCGACAGTAGCATCTTATCAGATTACTGACGAGTACGATGCTTACTTGGATATGGTAG 6222

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
              ||||  ||     ||| || ||||| || || || || || ||||| || |||||  ||||
Db       6223 ACGGGACAGTCGCCTGCCTGGATACTGCAACCTTCTGCCCCGCTAAGCTTAGAAGTTACC 6282

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
              | || || |||   |||     ||  | ||  |    || || |||||  | || ||||
Db       6283 CGAAAAAACATGAGTATAGAGCCCCGAATATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6342

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
              | ||  ||||||||||  |   ||| || |||||||| |||||||||| || || ||||
```

```
Db       6343 ACACGCTACAAAATGTGCTCATTGCCGCAACTAAAAGAAATTGCAACGTCACGCAGATGC 6402

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
              | ||| | ||     | || || ||| |||| ||||| || || ||    ||||| ||||
Db       6403 GTGAACTGCCAACACTGGACTCAGCGACATTCAATGTCGAATGCTTTCGAAAATATGCAT 6462

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
              | ||||| ||||| |||||     ||         || || ||| || | || || |||   |
Db       6463 GTAATGACGAGTATTGGGAGGAGTTCGCTCGGAAGCCAATTAGGATTACCACTGAGTTTG 6522

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
              | ||    ||||||  | |  |||||||| || || || ||  || | ||||| || ||
Db       6523 TCACCGCATATGTAGCTAGACTGAAAGGCCCTAAGGCCGCCACACTATTTGCAAAGACGT 6582

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
              |||||  |    || ||||| ||  | || ||||||  |||||||||||  | || ||||
Db       6583 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6642

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
              | || |||||||| || ||||||||||| ||||| ||  | || || || || || || | |
Db       6643 ACGTGAAAGTTACACCAGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6702

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
              | || ||||| ||||| || || |||||| | |||||||| || ||||||||||| ||||
Db       6703 AAGCCGCAGAACCCCTGGCGACTGCTTACTTATGCGGGATTCACCGGGAATTAGTGCGTA 6762

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
              | || |  || ||  | || ||||| || || || || || |||||||| ||||| ||||
Db       6763 GGCTTACGGCCGTCTTGCTTCCAAACATTCACACGCTTTTTGACATGTCGGCGGAGGATT 6822

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
              ||||||| || || || |||| | || |||||||| ||| |||| ||||| ||||
Db       6823 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGCGACCCGGTACTGGAGACGGATATCG 6882

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
              | || || || |||||| ||||||||||||| ||  |  | |  |||||||| | ||||
Db       6883 CATCATTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGTCTGATGATCTTGGAGG 6942

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
              || | ||||| || ||||| ||  | || ||||| |||   || || ||  | ||  |||
Db       6943 ACCTGGGTGTGGATCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 7002

Qy       1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
              |    || |||||||| || ||  | |||||||| || || ||||||||||||||| ||||
Db       7003 CCACCCATCTACCTACGGGTACTCGTTTTAAATTCGGGGCGATGATGAAATCCGGAATGT 7062

Qy       1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
              || | || || ||||||||||||  |  | ||| || | || || ||||||||||||  |
Db       7063 TCCTCACACTTTTTGTCAACACAGTTTTGAATGTCGTTATCGCCAGCAGAGTACTAGAAG 7122

Qy       1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
              | ||| | |  |||||     || || || |  || |||||||| |||||  | |||||  |
Db       7123 AGCGGCTTAAAACGTCCAGATGTGCAGCGTTCATTGGCGACGACAACATCATACATGGAG 7182

Qy       2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
              | || || ||||     ||||| ||||| |||||||| ||||| |||||||| || || |
Db       7183 TAGTATCTGACAAAGAAATGGCTGAGAGGTGCGCCACCTGGCTCAACATGGAGGTTAAGA 7242

Qy       2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139
              | || || ||||| || |||   | | |||| |||||||| || ||||||||| ||   |
Db       7243 TCATCGACGCAGTCATCGGTGAGAGACCACCTTACTTCTGCGGCGGATTTATCTTGCAAG 7302

Qy       2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199
              |   | | ||   |||||| ||| | || || || || || ||||||||| ||||||  | |
Db       7303 ATTCGGTTACTTCCACAGCGTGCCGCGTGGCGGATCCCCTGAAAAGGCTGTTTAAGTTGG 7362

Qy       2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
              | ||||| | |||| ||| ||  |||||||   ||| |  | || || ||||  ||||
Db       7363 GTAAACCGCTCCCAGCCGACGACGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 7422

Qy       2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319
              || ||| |   |||   ||| | || || || | |   ||||    ||||||      ||
Db       7423 AAACAAAGGCGTGGTTTAGAGTAGGTATAACAGGCACTTTAGCAGTGGCCGTGACGACCC 7482
```


Qy 2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379
Db 7483 GGTATGAGGTAGACAATATTACACCTGTCCTACTGGCATTGAGAACTTTTGCCCAGAGCA 7542
Qy 2380 TTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGG 2439
Db 7543 AAAGAGCATTCCAAGCCATCAGAGGGGAAATAAAGCATCTCTACGGTGGTCCTAAATAGT 7602
Qy 2440 TGACGTAGTAGACACGCACCTAC----------CCACCGGCAGAATGTTTCCATACC 2486
Db 7603 CAGCATAGTACATTTCATCTGACTAATACTACAACACCACCACCATGAATAGAGGATTCT 7662
Qy 2487 CTCAGCTGAACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACC 2546
Db 7663 TTAACATGCTCGGCCGCCGCCCCTTCCCGGCCCCCACTGCCATGTGGAGGCCGCGGAGAA 7722
Qy 2547 CTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTA 2606
Db 7723 GGAGGCAGGCGGCCCCGATGCCTGCCCGCAACGGGCTGGCTTCTCAAATCCAGCAACTGA 7782
Qy 2607 GGAGGTCGATAGTCAACTTGACTTTCAAACAACGATCACCTAATCCGCCGCCAGGTCCA- 2665
Db 7783 CCACAGCCGTCAGTGCCCTAGTCATTGGACAGGCAACTAGACCTCAACCCCCACGTCCAC 7842
Qy 2666 -----CCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGA 2720
Db 7843 GCCCGCCACCGCGCCAGAAGAAGCAGGCGCCCAAGCAACCACCGAAGCCGAAGAAACCAA 7902
Qy 2721 AGAAGCAGCAAGCCAAGAGGACGAAAC---GCAAGCCTAAACCAGGGAAACGACAACGTA 2777
Db 7903 AAACGCAGGAGAAGAAGAAGAAGCAACCTGCAAAACCCAAACCCGGAAAGAGACAGCGCA 7962
Qy 2778 TGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATCATG---CTGAACGGCCAAGTGA 2834
Db 7963 TGGCACTTAAGTTGGAGGCCGACAGATTGTTCGACGTCAAGAACGAGGACGGAGATGTCA 8022
Qy 2835 ATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACTCCACGTTGAAGGAAAAA 2894
Db 8023 TCGGGCACGCACTGGCCATGGAAGGAAAGGTAATGAAACCTCTGCACGTGAAAGGAACCA 8082
Qy 2895 TTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAGCATGTACGACTTGGAGT 2954
Db 8083 TCGACCACCCTGTGCTATCAAAGCTCAAATTTACCAAGTCGTCAGCATACGACATGGAGT 8142
Qy 2955 ACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTACACCAGCGACAAACCAC 3014
Db 8143 TCGCACAGTTGCCAGTCAACATGAGAAGTGAGGCATTCACCTACACCAGTGAACACCCCG 8202
Qy 3015 CGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAATGGGAGATTTACCGTAC 3074
Db 8203 AAGGATTCTATAACTGGCACCACGGAGCGGTGCAGTATAGTGGAGGTAGATTTACCATCC 8262
Qy 3075 CGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGACAACAGAGGCAGAG 3134
Db 8263 CTCGCGGAGTAGGAGGCAGAGGAGACAGCGGTCGTCCGATCATGGATAACTCCGGTCGGG 8322
Qy 3135 TTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGCGCTTTCAGTGGTCA 3194
Db 8323 TTGTCGCGATAGTCCTCGGTGGCGCTGATGAAGGAACACGAACTGCCCTTTCGGTCGTCA 8382
Qy 3195 CTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTCTGAACCGTGGT--- 3251
Db 8383 CCTGGAATAGTAAAGGGAAGACAATTAAGACGACCCCGGAAGGGACAGAAGAGTGGTCCG 8442
Qy 3252 ------CACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGACAAAC 3305
Db 8443 CAGCACCACTGGTCACGGCAATGTGTTTGCTCGGAAATGTGAGCTTCCCATGCGACCGCC 8502
Qy 3306 CACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAACGTCG 3365
Db 8503 CGCCCACATGCTATACCCGCGAACCTTCCAGAGCCCTCGACATCCTTGAAGAGAACGTGA 8562
Qy 3366 ACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAATGTCCA---TCACGCCGGC 3422
Db 8563 ACCATGAGGCCTACGATACCCTGCTCAATGCCATATTGCGGTGCGGATCGTCTGGCAGAA 8622
Qy 3423 CCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCCCGT 3482
Db 8623 GCAAAAGAAGCGTCGTTGACGACTTTACCCTGACCAGCCCCTACTTGGGCACATGCTCGT 8682
Qy 3483 ATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACGAAT 3542
Db 8683 ACTGCCACCATACTGAACCGTGCTTCAGCCCTGTTAAGATCGAGCAGGTCTGGGACGAAG 8742
Qy 3543 CTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAGGCA 3602
Db 8743 CGGACGATAACACCATACGCATACAGACTTCCGCCCAGTTTGGATACGACCAAAGCGGAG 8802
Qy 3603 CTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAAGGAAG 3662
Db 8803 CAGCAAGCGCAAACAAGTACCGCTACATGTCGCTTAAGCAGGATCACACCGTTAAAGAAG 8862
Qy 3663 ACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCACAAAG 3722
Db 8863 GCACCATGGATGACATCAAGATTAGCACCTCAGGACCGTGTAGAAGGCTTAGCTACAAAG 8922
Qy 3723 GGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCACGAGCG 3782
Db 8923 GATACTTTCTCCTCGCAAAATGCCCTCCAGGGGACAGCGTAACGGTTAGCATAGTGAGTA 8982
Qy 3783 GAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTAGAG 3842
Db 8983 GCAACTCAGCAACGTCATGTACACTGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGG 9042
Qy 3843 AGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATCACT 3902
Db 9043 AAAAATATGATCTACCTCCCGTTCACGGTAAAAGAATTCCTTGCACAGTGTACGACCGTC 9102
Qy 3903 TGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAAGT 3962
Db 9103 TGAAA---ACAACTGCAGGCTACATCACTATGCACAGGCCGGGACCGCACGCTTATACAT 9159
Qy 3963 CCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGTCA 4022
Db 9160 CCTACCTGGAAGAATCATCAGGGAAAGTTTACGCAAAGCCGCCATCTGGGAAGAACATTA 9219
Qy 4023 CCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGATGA 4082
Db 9220 CGTATGAGTGCAAGTGCGGCGACTACAAGACCGGAACCGTTTCGACCCGCACCGAAATCA 9279
Qy 4083 ACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGTCT 4142
Db 9280 CTGGTTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAACGAAGTGGGTCT 9339
Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9340 TCAACTCACCGGACTTGATCAGACATGACGACCACACGGCCCAAGGGAAATTGCATTTGC 9399
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9400 CTTTCAAGTTGATCCCGGGTGCCTGCATGGTCCCTGTTGCCCACGCGCCGAATGTAATAC 9459
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9460 ATGGCTTTAAACACATCAGCCTCCAATTAGATACAGACCACTTGACATTGCTCACCACCA 9519
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9520 GGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCGTCGGAAAGACGGTCAGAA 9579
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9580 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGAAATCATGAGCCAGTGAGGG 9639
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502

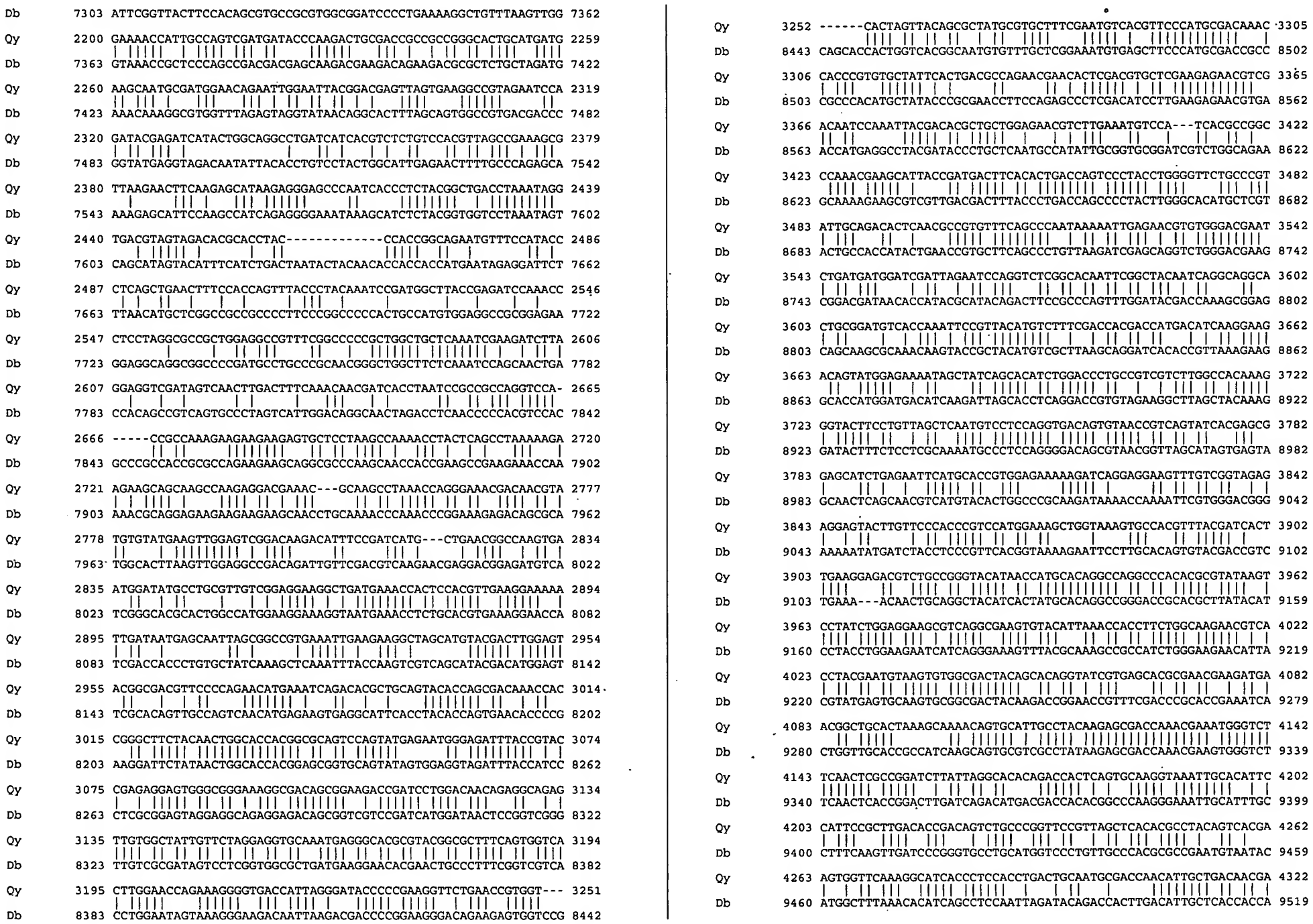
```
Db       9640 TCTATGCCCAAGAGTCAGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9699

Qy       4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562

Db       9700 ATTACTACCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTACCGTGGCGA 9759

Qy       4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622

Db       9760 TGATGATTGGCGTAACTGTTGCAGTGTTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9819

Qy       4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682

Db       9820 CGCCATACGCCCTGGCCCCAAACGCCGTAATCCCAACTTCGCTGGCACTCTTGTGCTGCG 9879

Qy       4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742

Db       9880 TTAGGTCGGCCAATGCTGAAACGTTCACCGAGACCATGAGTTACTTGTGGTCGAACAGTC 9939

Qy       4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802

Db       9940 AGCCGTTCTTCTGGGTCCAGTTGTGCATACCTTTGGCCGCGTTCATCGTTCTAATGCGCT 9999

Qy       4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862

Db      10000 ACTGCTCCTGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10059

Qy       4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922

Db      10060 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10119

Qy       4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982

Db      10120 AAAGGGCAGGGTATGCCCCGCTCAATTTGGAGATCACTGTCATGTCCTCGGAGGTTTTGC 10179

Qy       4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042

Db      10180 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCAAAAA 10239

Qy       5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102

Db      10240 TCAAATGCTGCGGCTCCTTGGAATGTCAGCCGGCCGCTCATGCAGACTATACCTGCAAGG 10299

Qy       5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162

Db      10300 TCTTCGGAGGGGTCTACCCCTTTATGTGGGGAGGAGCGCAATGTTTTTGCGACAGTGAGA 10359

Qy       5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222

Db      10360 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGCAGATTGCGCGTCTGACCACGCGC 10419

Qy       5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282

Db      10420 AGGCGATTAAGGTGCACACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10479

Qy       5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342

Db      10480 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACTTGA 10539

Qy       5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402

Db      10540 AAGTCATAGCTGGACCAATTTCAGCATCGTTTACGCCATTCGATCATAAGGTCGTTATCC 10599

Qy       5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462

Db      10600 ATCGCGGCCTGGTGTACAACTATGACTTCCCGGAATATGGAGCGATGAAACCAGGAGCGT 10659

Qy       5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522

Db      10660 TCGGAGACATTCAAGCTACCTCCTTGACTAGCAAGGATCTCATCGCCAGCACAGACATTA 10719

Qy       5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582

Db      10720 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCAGGAT 10779

Qy       5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642

Db      10780 TTGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCACCTTTCGGGTGTA 10839

Qy       5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702

Db      10840 AGATTGCAGTAAATCCGCTCCGAGCGGTGGACTGTTCATACGGGAACATTCCCATTTCTA 10899

Qy       5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762

Db      10900 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10959

Qy       5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822

Db      10960 AATGTGAAGTCAGTGAGTGCACTTATTCAGCAGACTTCGGCGGGATGGCCACCCTGCAGT 11019

Qy       5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882

Db      11020 ATGTATCCGACCGCGAAGGTCAATGCCCCGTACATTCGCATTCGAGCACAGCAACTCTCC 11079

Qy       5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942

Db      11080 AAGAGTCGACAGTACATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTTAGCACCGCGA 11139

Qy       5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002

Db      11140 GTCCACAGGCGAACTTTATCGTATCGCTGTGTGGGAAGAAGACAACATGCAATGCAGAAT 11199

Qy       6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062

Db      11200 GTAAACCACCAGCTGACCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTTCAAG 11259

Qy       6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122

Db      11260 CCGCCATCTCAAAAACATCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11319

Qy       6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182

Db      11320 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11379

Qy       6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241

Db      11380 GACCGCTACGCCCCAATGATCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11439

Qy       6242 GCATAATGCCACGCGCCGCT 6261

Db      11440 GCATAATGCATCAGGCTGGT 11459
```

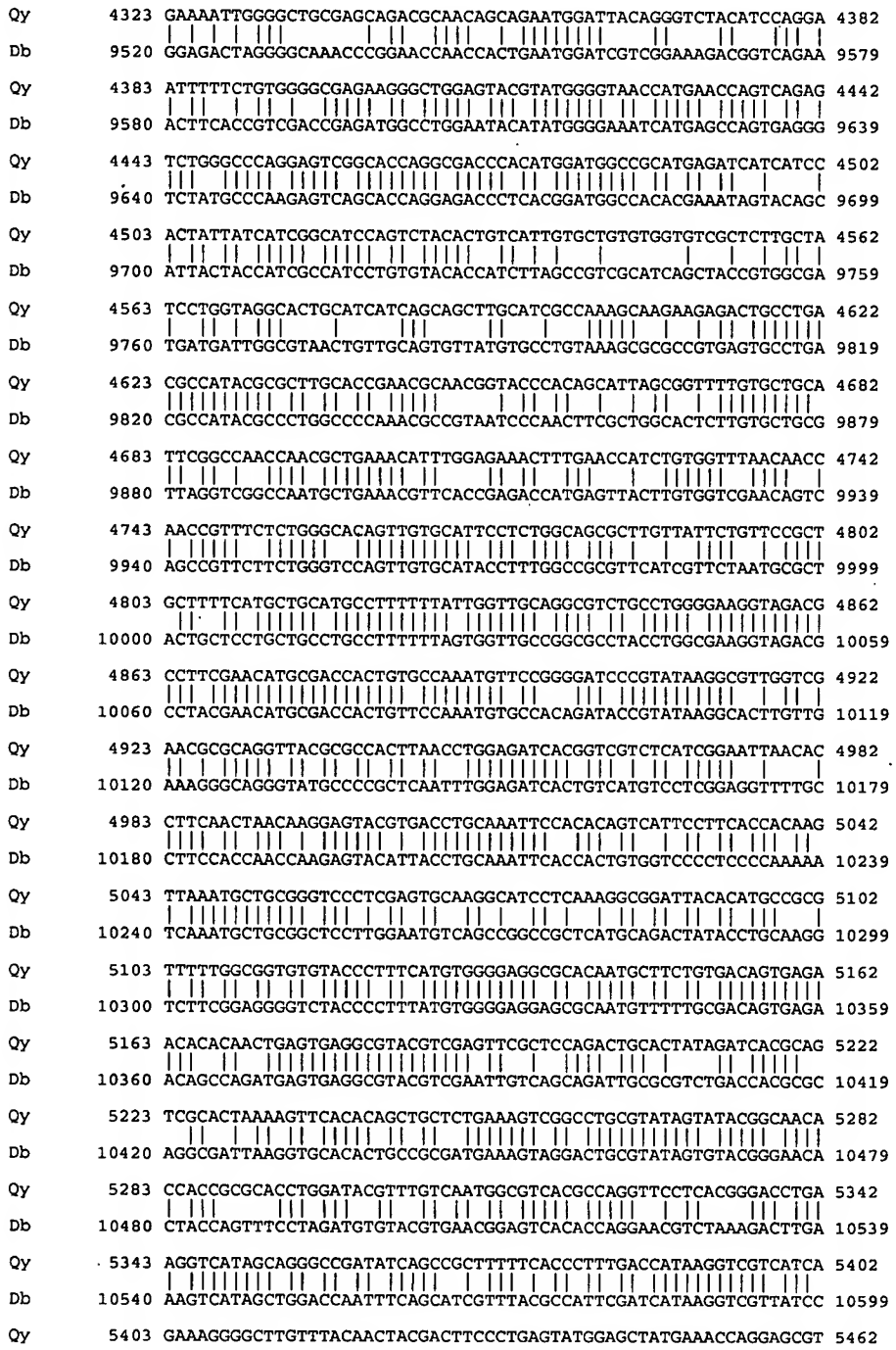
```
RESULT 14
AAV70684
ID   AAV70684 standard; DNA; 16656 BP.
XX
AC   AAV70684;
XX
DT   20-MAR-2003  (updated)
DT   02-FEB-1999  (first entry)
XX
DE   Representative eukaryotic layered vector initiation system.
XX
KW   Eukaryotic layered vector initiation system;
KW   alphavirus vector construct; gene therapy; ss.
XX
OS   Synthetic.
OS   Sindbis virus.
XX
PN   US5843723-A.
XX
PD   01-DEC-1998.
XX
PF   30-OCT-1996;   96US-0739167.
```

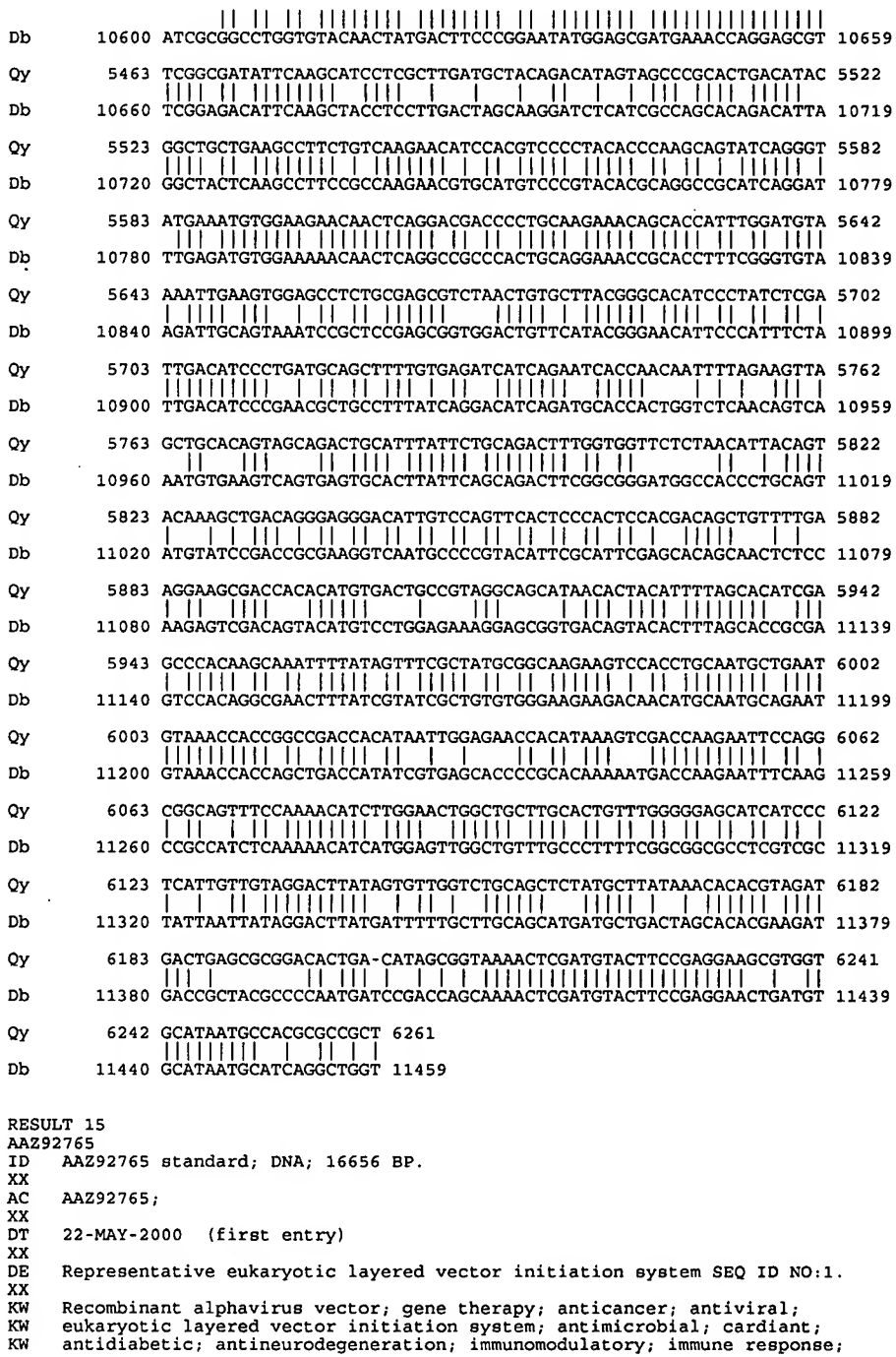
XX
PR 15-MAR-1995; 95US-0404796.
PR 15-SEP-1993; 93US-0122791.
PR 18-FEB-1994; 94US-0198450.
PR 30-NOV-1994; 94US-0348472.
PR 20-JAN-1995; 95US-0376184.
XX
PA (CHIR) CHIRON CORP.
XX
PI Belli BA, Chang SMW, Driver DA, Dubensky TW, Ibanez CE;
PI Jolly DJ, Polo JM;
XX
DR WPI; 1999-044581/04.
XX
PT Alphavirus vectors constructs containing a 5' promoter of viral cDNA
PT by in vitro transcription - used in gene therapy
XX
PS Disclosure; Fig 3A-H; 140pp; English.
XX
CC The present sequence represents a representative eukaryotic layered
CC vector initiation system derived from Sindbis. The sequence exemplifies
CC the invention. The specification describes alphavirus vector constructs.
CC These constructs comprise a promoter 5' of viral cDNA which initiates
CC the synthesis of RNA from the viral cDNA by in vitro transcription,
CC followed by a 5' sequence which initiates transcription of alphavirus
CC RNA, followed by a nucleotide sequence encoding alphavirus nonstructural
CC proteins, a viral junction region which has been inactivated such that
CC viral transcription of a subgenomic fragment is prevented, an internal
CC ribosome entry site or a sequence which promotes ribosome read through
CC between adjacent reading frames, and an alphavirus RNA polymerase
CC recognition sequence. The recombinant alphavirus vectors can be used
CC for gene therapy.
CC (Updated on 20-MAR-2003 to correct PR field.)
XX
SQ Sequence 16656 BP; 4467 A; 4349 C; 4183 G; 3657 T; 0 other;

Query Match 36.5%; Score 2369.2; DB 20; Length 16656;
Best Local Similarity 64.6%; Pred. No. 0;
Matches 3693; Conservative 0; Mismatches 1983; Indels 44; Gaps 9;

```
Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
              ||||||     |  |  || | ||| || || ||  |||| |||||||  || ||| | |
Db       5746 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5805

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
              || |||| || || |  ||     | ||   |||||| |  ||||| ||    ||||  |
Db       5806 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTCCTGG 5865

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
              | |    | | ||||||   |||||        | |||| ||    |  |    |     |
Db       5866 AAAGAATTCATGCCCCGGTGCTCGACACGTCGAAAGAGGAACAACTCAAACTCAGGTACC 5925

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
              |  |      || | ||||  || | ||| |||||  || || || ||||||||||||  ||
Db       5926 AGATGATGCCCACCGAAGCCAACAAAAGTAGGTACCAGTCTCGTAAAGTAGAAAATCAGA 5985

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
              |||| || ||  | |||||||||  | || ||| |  |    |||   ||  |   ||
Db       5986 AAGCCATAACCACTGAGCGACTACTGTCAGGACTACGACTGTATAACTCTGCCACAG--- 6042

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
              |||      || || |  |||  |||||   |||| | |||||     |  |||
Db       6043 ATCAGCCAGAATGCTATAAGATCACCTATCCGAAACCATTGTACTCCAGTAGCGTACCGG 6102

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
                |  |   |  | || || |||| ||      || |||    | | || |||||  |
Db       6103 CGAACTACTCCGATCCACAGTTCGCTGTAGCTGTCTGTAACAACTATCTGCATGAGAACT 6162

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
              | || ||||||||   ||||  || || || || |||||||| ||  | || |||||  |
Db       6163 ATCCGACAGTAGCATCTTATCAGATTACTGACGAGTACGATGCTTACTTGGATATGGTAG 6222

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
              |||| ||     ||| || ||||| || || || || || ||||| || ||||| ||||
Db       6223 ACGGGACAGTCGCCTGCCTGGATACTGCAACCTTCTGCCCCGCTAAGCTTAGAAGTTACC 6282

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
              | || || |||   |||       ||  | ||  |    || || |||||  | || ||||
Db       6283 CGAAAAAACATGAGTATAGAGCCCCGAATATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6342

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
              | ||  |||||||||  |   ||| || |||||||| |||||||||| || || ||||
Db       6343 ACACGCTACAAAATGTGCTCATTGCCGCAACTAAAAGAAATTGCAACGTCACGCAGATGC 6402

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
              | ||| | ||      | || || ||| |||| ||||| || || ||    ||||| ||||
Db       6403 GTGAACTGCCAACACTGGACTCAGCGACATTCAATGTCGAATGCTTTCGAAAATATGCAT 6462

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
              | ||||| |||||  |||||     ||         || || ||| || | || ||  |
Db       6463 GTAATGACGAGTATTGGGAGGAGTTCGCTCGGAAGCCAATTAGGATTACCACTGAGTTTG 6522

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
              | ||   ||||||  | |  |||||||| || || || ||  || | ||||| || ||
Db       6523 TCACCGCATATGTAGCTAGACTGAAAGGCCCTAAGGCCGCCACACTATTTGCAAAGACGT 6582

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
              |||||  |    || ||||| ||  | || ||||||  |||||||||||  | || ||||
Db       6583 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6642

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
              | || |||||||| || |||||||||| ||||| ||  | || || || || || || |
Db       6643 ACGTGAAAGTTACACCAGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6702

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
              | || ||||| ||||| || || |||||| | |||||||| || ||||||||||| ||||
Db       6703 AAGCCGCAGAACCCCTGGCGACTGCTTACTTATGCGGGATTCACCGGGAATTAGTGCGTA 6762

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
              | || |  || ||  | || ||||| || || || || || |||||||| ||||| ||||
Db       6763 GGCTTACGGCCGTCTTGCTTCCAAACATTCACACGCTTTTTGACATGTCGGCGGAGGATT 6822

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
              |||||| || || || ||||| ||| | || |||||||| ||| |||| ||||| ||||
Db       6823 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGCGACCCGGTACTGGAGACGGATATCG 6882

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
              | || || || |||||| |||||||||||||| ||  |  | |  |||||||| | ||||
Db       6883 CATCATTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGTCTGATGATCTTGGAGG 6942

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
              || | |||||  || |||||  ||  | ||  |||||  |||   ||  ||  ||  | ||  |||
Db       6943 ACCTGGGTGTGGATCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 7002

Qy       1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
              |     || |||||||| || ||  | |||||||  || || ||||||||||||||  ||||
Db       7003 CCACCCATCTACCTACGGGTACTCGTTTTAAATTCGGGGCGATGATGAAATCCGGAATGT 7062

Qy       1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
              || | || || |||||||||||| |  | ||| || | || || ||||||||||||  |
Db       7063 TCCTCACACTTTTTGTCAACACAGTTTTGAATGTCGTTATCGCCAGCAGAGTACTAGAAG 7122

Qy       1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
              | ||| | |  |||||     || || || |  || |||||||| |||||  | |||||
Db       7123 AGCGGCTTAAAACGTCCAGATGTGCAGCGTTCATTGGCGACGACAACATCATACATGGAG 7182

Qy       2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
              | || || ||||     ||||| ||||| |||||||| |||| |||||||| || || |
Db       7183 TAGTATCTGACAAAGAAATGGCTGAGAGGTGCGCCACCTGGCTCAACATGGAGGTTAAGA 7242

Qy       2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139
              | || || ||||| || |||   | | |||| |||||P|| || |||||||||  ||  |
Db       7243 TCATCGACGCAGTCATCGGTGAGAGACCACCTTACTTCTGCGGCGGATTTATCTTGCAAG 7302

Qy       2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199
              |   | | ||  |||||| ||| | || || || || || |||||||| |||||| | |
```



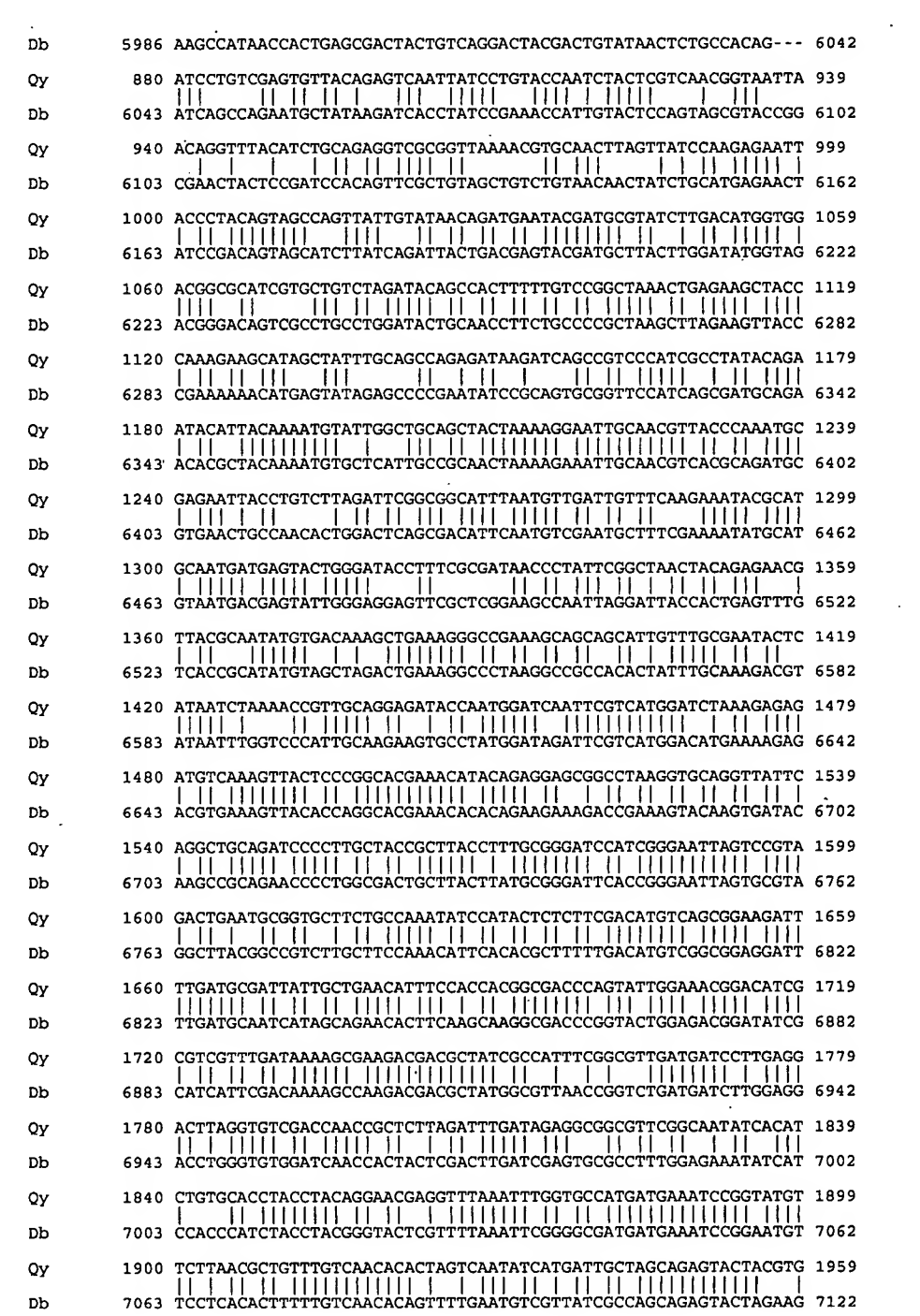


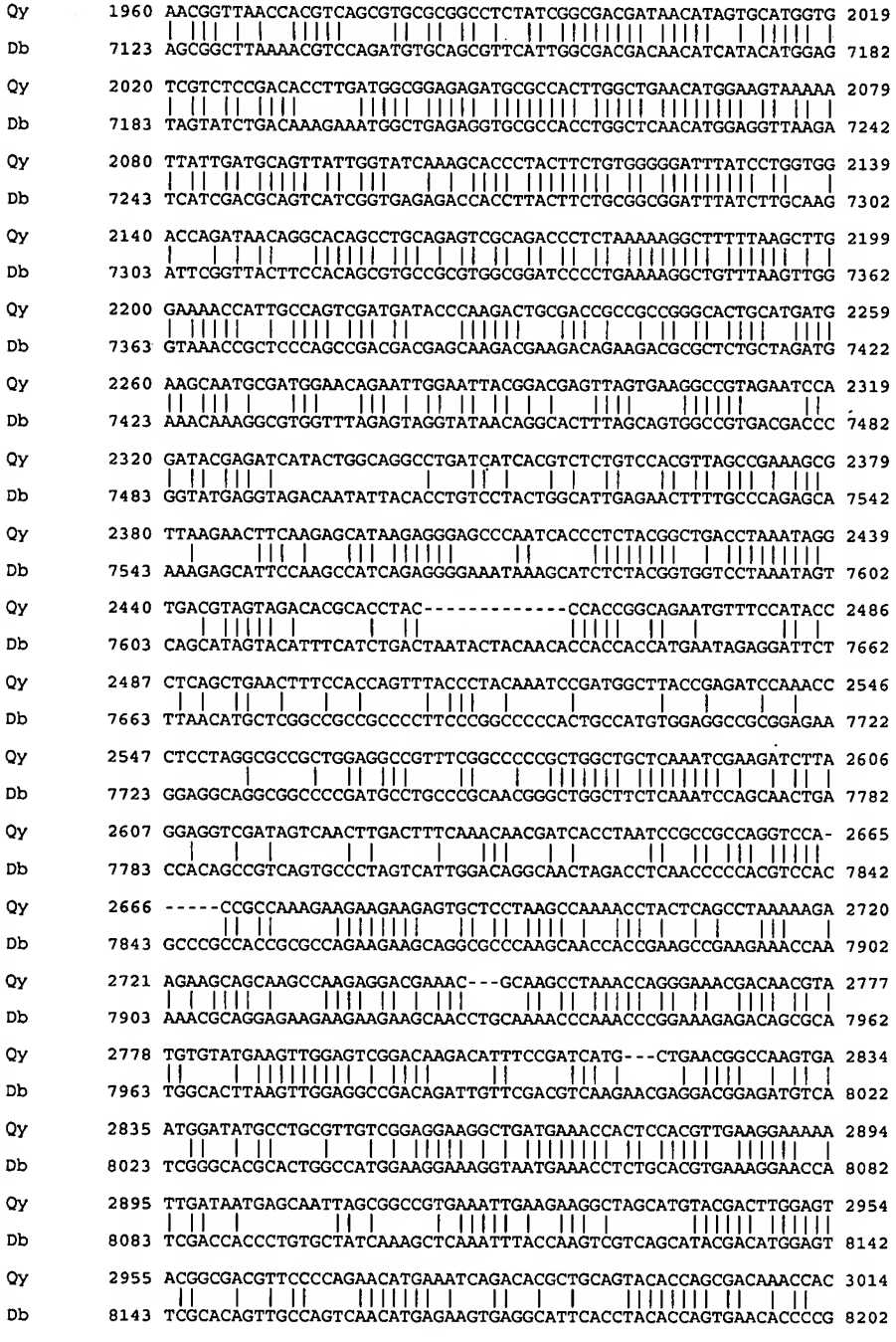


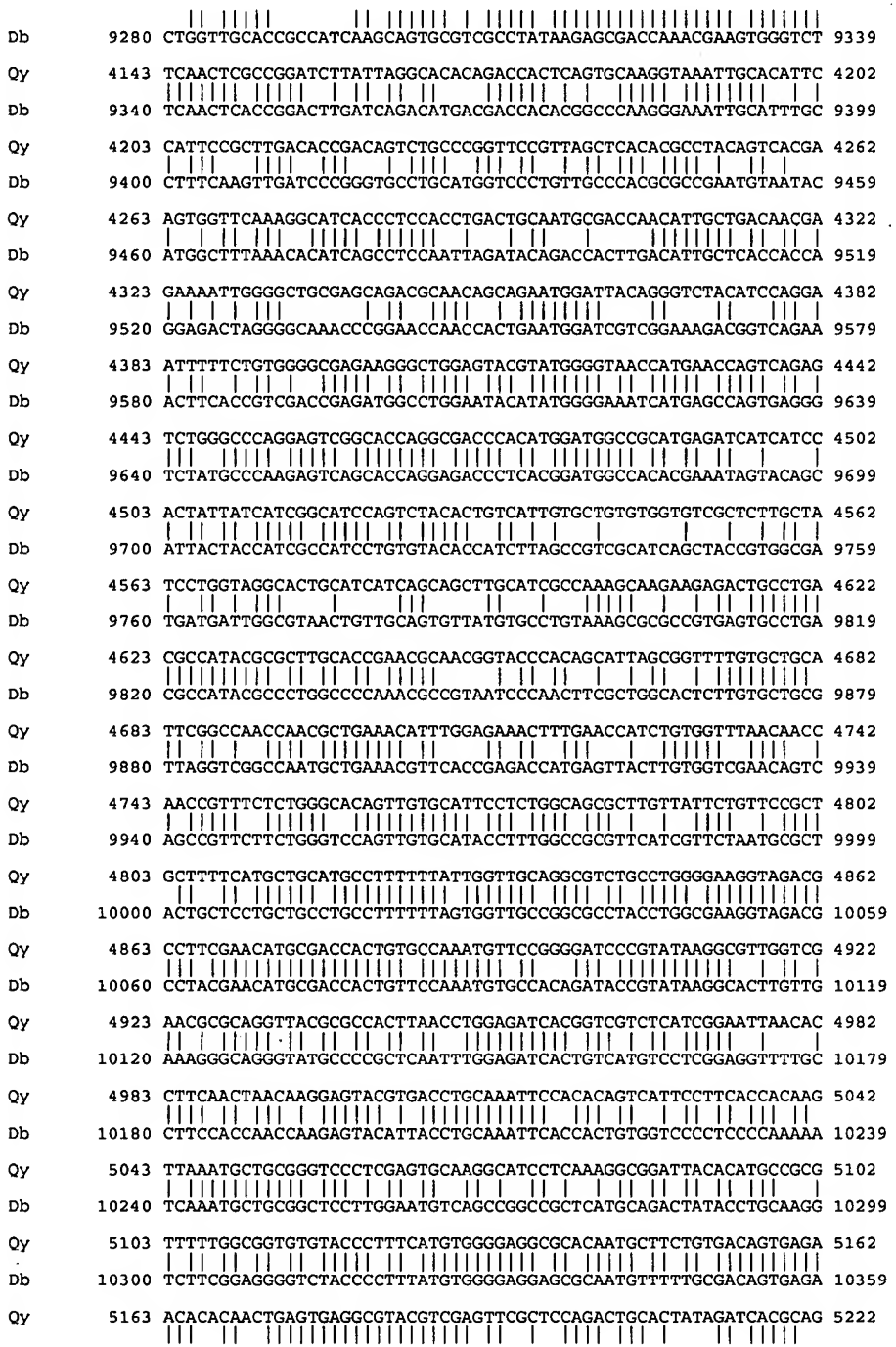
RESULT 15
AAZ92765
ID AAZ92765 standard; DNA; 16656 BP.
XX
AC AAZ92765;
XX
DT 22-MAY-2000 (first entry)
XX
DE Representative eukaryotic layered vector initiation system SEQ ID NO:1.
XX
KW Recombinant alphavirus vector; gene therapy; anticancer; antiviral;
KW eukaryotic layered vector initiation system; antimicrobial; cardiant;
KW antidiabetic; antineurodegeneration; immunomodulatory; immune response;

KW infection; melanoma; cancer; diabetes; autoimmune disorder;
KW graft versus host disease; Alzheimer's disease; heart disease;
KW haemophilia; cystic fibrosis; PCR primer; ss.
XX
OS Sindbis virus.
XX
PN US6015686-A.
XX
PD 18-JAN-2000.
XX
PF 15-MAR-1995; 95US-0404796.
XX
PR 15-SEP-1993; 93US-0122791.
PR 18-FEB-1994; 94US-0198450.
PR 30-NOV-1994; 94US-0348472.
PR 20-JAN-1995; 95US-0376184.
XX
PA (CHIR) CHIRON VIAGENE INC.
XX
PI Polo JM, Jolly DJ, Driver DA, Dubensky TW;
XX
DR WPI; 2000-181143/16.
XX
PT Eukaryotic layered vector initiation system useful for gene therapy and
PT production of recombinant protein, comprises promoter that directs
PT synthesis of RNA containing a vector construct -
XX
PS Example 2; Fig 3; 141pp; English.
XX
CC The present invention describes a eukaryotic layered vector initiation
CC system (A) comprising a eukaryotic promoter (EP), 5' of viral cDNA (I)
CC which initiates the 5' to 3' synthesis of RNA (II) from (I). (II)
CC comprises a vector construct (VC), expressing a heterologous nucleic
CC acid (III), which amplifies autonomously in a cell. (A) can have
CC anticancer, antiviral, antimicrobial, antidiabetic, immunomodulatory,
CC antineurodegeneration and cardiant activities. (A) are used to express
CC therapeutic proteins in cell cultures; in gene therapy (for humans or
CC animals), e.g. to induce a specific immune response; to inhibit
CC interaction of an agent with cellular receptors; to express a toxin;
CC to regulate the immune system or to express a replacement gene, e.g.
CC for treatment or prevention of infections (by viruses or other
CC pathogens), melanoma (or other cancers), diabetes (or other autoimmune
CC disorders), graft versus host disease, Alzheimer's disease, heart
CC disease, haemophilia, cystic fibrosis and many others; or for production
CC of packaged vector particles (also useful for gene therapy). (A) can
CC also be used to produce transgenic plants that express resistance or
CC growth promoting sequences. AAZ92765 to AAZ92891 and AAY80146 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 16656 BP; 4467 A; 4349 C; 4183 G; 3657 T; 0 other;

Query Match 36.5%; Score 2369.2; DB 21; Length 16656;
Best Local Similarity 64.6%; Pred. No. 0;
Matches 3693; Conservative 0; Mismatches 1983; Indels 44; Gaps 9;




Qy 1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
Db 7123 AGCGGCTTAAAACGTCCAGATGTGCAGCGTTCATTGGCGACGACAACATCATACATGGAG 7182
Qy 2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
Db 7183 TAGTATCTGACAAAGAAATGGCTGAGAGGTGCGCCACCTGGCTCAACATGGAGGTTAAGA 7242
Qy 2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139
Db 7243 TCATCGACGCAGTCATCGGTGAGAGACCACCTTACTTCTGCGGCGGATTTATCTTGCAAG 7302
Qy 2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199
Db 7303 ATTCGGTTACTTCCACAGCGTGCCGCGTGGCGGATCCCCTGAAAAGGCTGTTTAAGTTGG 7362
Qy 2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
Db 7363 GTAAACCGCTCCCAGCCGACGACGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 7422
Qy 2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319
Db 7423 AAACAAAGGCGTGGTTTAGAGTAGGTATAACAGGCACTTTAGCAGTGGCCGTGACGACCC 7482
Qy 2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379
Db 7483 GGTATGAGGTAGACAATATTACACCTGTCCTACTGGCATTGAGAACTTTTGCCCAGAGCA 7542
Qy 2380 TTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGG 2439
Db 7543 AAAGAGCATTCCAAGCCATCAGAGGGGAAATAAAGCATCTCTACGGTGGTCCTAAATAGT 7602
Qy 2440 TGACGTAGTAGACACGCACCTAC----------CCACCGGCAGAATGTTTCCATACC 2486
Db 7603 CAGCATAGTACATTTCATCTGACTAATACTACAACACCACCACCATGAATAGAGGATTCT 7662
Qy 2487 CTCAGCTGAACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACC 2546
Db 7663 TTAACATGCTCGGCCGCCGCCCCTTCCCGGCCCCCACTGCCATGTGGAGGCCGCGGAGAA 7722
Qy 2547 CTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTA 2606
Db 7723 GGAGGCAGGCGGCCCCGATGCCTGCCCGCAACGGGCTGGCTTCTCAAATCCAGCAACTGA 7782
Qy 2607 GGAGGTCGATAGTCAACTTGACTTTCAAACAACGATCACCTAATCCGCCGCCAGGTCCA- 2665
Db 7783 CCACAGCCGTCAGTGCCCTAGTCATTGGACAGGCAACTAGACCTCAACCCCCACGTCCAC 7842
Qy 2666 -----CCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGA 2720
Db 7843 GCCCGCCACCGCGCCAGAAGAAGCAGGCGCCCAAGCAACCACCGAAGCCGAAGAAACCAA 7902
Qy 2721 AGAAGCAGCAAGCCAAGAGGACGAAAC---GCAAGCCTAAACCAGGGAAACGACAACGTA 2777
Db 7903 AAACGCAGGAGAAGAAGAAGAAGCAACCTGCAAAACCCAAACCCGGAAAGAGACAGCGCA 7962
Qy 2778 TGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATCATG---CTGAACGGCCAAGTGA 2834
Db 7963 TGGCACTTAAGTTGGAGGCCGACAGATTGTTCGACGTCAAGAACGAGGACGGAGATGTCA 8022
Qy 2835 ATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACTCCACGTTGAAGGAAAAA 2894
Db 8023 TCGGGCACGCACTGGCCATGGAAGGAAAGGTAATGAAACCTCTGCACGTGAAAGGAACCA 8082
Qy 2895 TTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAGCATGTACGACTTGGAGT 2954
Db 8083 TCGACCACCCTGTGCTATCAAAGCTCAAATTTACCAAGTCGTCAGCATACGACATGGAGT 8142
Qy 2955 ACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTACACCAGCGACAAACCAC 3014
Db 8143 TCGCACAGTTGCCAGTCAACATGAGAAGTGAGGCATTCACCTACACCAGTGAACACCCCG 8202
Qy 3015 CGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAATGGGAGATTTACCGTAC 3074
Db 8203 AAGGATTCTATAACTGGCACCACGGAGCGGTGCAGTATAGTGGAGGTAGATTTACCATCC 8262
Qy 3075 CGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGACAACAGAGGCAGAG 3134
Db 8263 CTCGCGGAGTAGGAGGCAGAGGAGACAGCGGTCGTCCGATCATGGATAACTCCGGTCGGG 8322
Qy 3135 TTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGCGCTTTCAGTGGTCA 3194
Db 8323 TTGTCGCGATAGTCCTCGGTGGCGCTGATGAAGGAACACGAACTGCCCTTTCGGTCGTCA 8382
Qy 3195 CTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTCTGAACCGTGGT--- 3251
Db 8383 CCTGGAATAGTAAAGGGAAGACAATTAAGACGACCCCGGAAGGGACAGAAGAGTGGTCCG 8442
Qy 3252 ------CACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGACAAAC 3305
Db 8443 CAGCACCACTGGTCACGGCAATGTGTTTGCTCGGAAATGTGAGCTTCCCATGCGACCGCC 8502
Qy 3306 CACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAACGTCG 3365
Db 8503 CGCCCACATGCTATACCCGCGAACCTTCCAGAGCCCTCGACATCCTTGAAGAGAACGTGA 8562
Qy 3366 ACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAATGTCCA---TCACGCCGGC 3422
Db 8563 ACCATGAGGCCTACGATACCCTGCTCAATGCCATATTGCGGTGCGGATCGTCTGGCAGAA 8622
Qy 3423 CCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCCCGT 3482
Db 8623 GCAAAAGAAGCGTCGTTGACGACTTTACCCTGACCAGCCCCTACTTGGGCACATGCTCGT 8682
Qy 3483 ATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACGAAT 3542
Db 8683 ACTGCCACCATACTGAACCGTGCTTCAGCCCTGTTAAGATCGAGCAGGTCTGGGACGAAG 8742
Qy 3543 CTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAGGCA 3602
Db 8743 CGGACGATAACACCATACGCATACAGACTTCCGCCCAGTTTGGATACGACCAAAGCGGAG 8802
Qy 3603 CTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAAGGAAG 3662
Db 8803 CAGCAAGCGCAAACAAGTACCGCTACATGTCGCTTAAGCAGGATCACACCGTTAAAGAAG 8862
Qy 3663 ACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCACAAAG 3722
Db 8863 GCACCATGGATGACATCAAGATTAGCACCTCAGGACCGTGTAGAAGGCTTAGCTACAAAG 8922
Qy 3723 GGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCACGAGCG 3782
Db 8923 GATACTTTCTCCTCGCAAAATGCCCTCCAGGGGACAGCGTAACGGTTAGCATAGTGAGTA 8982
Qy 3783 GAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTAGAG 3842
Db 8983 GCAACTCAGCAACGTCATGTACACTGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGG 9042
Qy 3843 AGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATCACT 3902
Db 9043 AAAAATATGATCTACCTCCCGTTCACGGTAAAAGAATTCCTTGCACAGTGTACGACCGTC 9102
Qy 3903 TGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAAGT 3962
Db 9103 TGAAA---ACAACTGCAGGCTACATCACTATGCACAGGCCGGGACCGCACGCTTATACAT 9159
Qy 3963 CCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGTCA 4022
Db 9160 CCTACCTGGAAGAATCATCAGGGAAAGTTTACGCAAAGCCGCCATCTGGGAAGAACATTA 9219
Qy 4023 CCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGATGA 4082
Db 9220 CGTATGAGTGCAAGTGCGGCGACTACAAGACCGGAACCGTTTCGACCCGCACCGAAATCA 9279
Qy 4083 ACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGTCT 4142


Db 9280 CTGGTTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAACGAAGTGGGTCT 9339
Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9340 TCAACTCACCGGACTTGATCAGACATGACGACCACACGGCCCAAGGGAAATTGCATTTGC 9399
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9400 CTTTCAAGTTGATCCCGGGTGCCTGCATGGTCCCTGTTGCCCACGCGCCGAATGTAATAC 9459
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9460 ATGGCTTTAAACACATCAGCCTCCAATTAGATACAGACCACTTGACATTGCTCACCACCA 9519
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9520 GGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCGTCGGAAAGACGGTCAGAA 9579
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9580 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGAAATCATGAGCCAGTGAGGG 9639
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db 9640 TCTATGCCCAAGAGTCAGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9699
Qy 4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db 9700 ATTACTACCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTACCGTGGCGA 9759
Qy 4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db 9760 TGATGATTGGCGTAACTGTTGCAGTGTTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9819
Qy 4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db 9820 CGCCATACGCCCTGGCCCCAAACGCCGTAATCCCAACTTCGCTGGCACTCTTGTGCTGCG 9879
Qy 4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db 9880 TTAGGTCGGCCAATGCTGAAACGTTCACCGAGACCATGAGTTACTTGTGGTCGAACAGTC 9939
Qy 4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
Db 9940 AGCCGTTCTTCTGGGTCCAGTTGTGCATACCTTTGGCCGCGTTCATCGTTCTAATGCGCT 9999
Qy 4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
Db 10000 ACTGCTCCTGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10059
Qy 4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
Db 10060 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10119
Qy 4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db 10120 AAAGGGCAGGGTATGCCCCGCTCAATTTGGAGATCACTGTCATGTCCTCGGAGGTTTTGC 10179
Qy 4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db 10180 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCAAAAA 10239
Qy 5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db 10240 TCAAATGCTGCGGCTCCTTGGAATGTCAGCCGGCCGCTCATGCAGACTATACCTGCAAGG 10299
Qy 5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db 10300 TCTTCGGAGGGGTCTACCCCTTTATGTGGGGAGGAGCGCAATGTTTTTGCGACAGTGAGA 10359
Qy 5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db 10360 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGCAGATTGCGCGTCTGACCACGCGC 10419
Qy 5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db 10420 AGGCGATTAAGGTGCACACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10479
Qy 5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db 10480 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACTTGA 10539
Qy 5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db 10540 AAGTCATAGCTGGACCAATTTCAGCATCGTTTACGCCATTCGATCATAAGGTCGTTATCC 10599
Qy 5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db 10600 ATCGCGGCCTGGTGTACAACTATGACTTCCCGGAATATGGAGCGATGAAACCAGGAGCGT 10659
Qy 5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db 10660 TCGGAGACATTCAAGCTACCTCCTTGACTAGCAAGGATCTCATCGCCAGCACAGACATTA 10719
Qy 5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db 10720 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCAGGAT 10779
Qy 5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db 10780 TTGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCACCTTTCGGGTGTA 10839
Qy 5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db 10840 AGATTGCAGTAAATCCGCTCCGAGCGGTGGACTGTTCATACGGGAACATTCCCATTTCTA 10899
Qy 5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db 10900 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10959
Qy 5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db 10960 AATGTGAAGTCAGTGAGTGCACTTATTCAGCAGACTTCGGCGGGATGGCCACCCTGCAGT 11019
Qy 5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db 11020 ATGTATCCGACCGCGAAGGTCAATGCCCCGTACATTCGCATTCGAGCACAGCAACTCTCC 11079
Qy 5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db 11080 AAGAGTCGACAGTACATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTTAGCACCGCGA 11139
Qy 5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db 11140 GTCCACAGGCGAACTTTATCGTATCGCTGTGTGGGAAGAAGACAACATGCAATGCAGAAT 11199
Qy 6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db 11200 GTAAACCACCAGCTGACCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTTCAAG 11259
Qy 6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db 11260 CCGCCATCTCAAAAACATCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11319
Qy 6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
Db 11320 TATTAATTATAGGACTTATGATTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11379
Qy 6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
Db 11380 GACCGCTACGCCCCAATGATCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11439
Qy 6242 GCATAATGCCACGCGCCGCT 6261
Db 11440 GCATAATGCATCAGGCTGGT 11459

Search completed: November 15, 2003, 23:18:21
Job time : 1555 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2003, 05:39:26 ; Search time 1783 Seconds
(without alignments)
11886.879 Million cell updates/sec

Title: US-10-023-649-1_COPY_5000_11484
Perfect score: 6485
Sequence: 1 caccgccagcaacgacaacc.....tattttgtttttaaaatttc 6485

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6485	100.0	11484	12	US-10-023-649-1	Sequence 1, Appli
2	4137.2	63.8	4395	12	US-10-023-649-7	Sequence 7, Appli
3	4129.8	63.7	4150	12	US-10-023-649-5	Sequence 5, Appli
4	2380.4	36.7	11740	11	US-09-507-362-103	Sequence 103, App
5	1442.6	22.2	11517	10	US-09-901-106-1	Sequence 1, Appli
6	926.2	14.3	12523	10	US-09-991-258-1	Sequence 1, Appli
7	924.6	14.3	12379	10	US-09-991-258-14	Sequence 14, Appl
8	924.6	14.3	13584	10	US-09-991-258-17	Sequence 17, Appl
9	916.2	14.1	7479	10	US-09-991-258-2	Sequence 2, Appli
10	827.6	12.8	8000	11	US-09-507-362-102	Sequence 102, App
11	824.4	12.7	8000	11	US-09-507-362-101	Sequence 101, App
12	816	12.6	11282	10	US-09-733-042-1	Sequence 1, Appli
13	816	12.6	11282	11	US-09-275-883-1	Sequence 1, Appli
14	764	11.8	8100	12	US-09-190-246-4	Sequence 4, Appli
15	764	11.8	10610	11	US-09-994-412-3	Sequence 3, Appli
16	764	11.8	15538	12	US-09-190-246-1	Sequence 1, Appli
17	607.6	9.4	2943	10	US-09-991-258-12	Sequence 12, Appl
18	607.6	9.4	6989	10	US-09-991-258-11	Sequence 11, Appl
19	307.4	4.7	1026	10	US-09-991-258-9	Sequence 9, Appli
20	307.4	4.7	5076	10	US-09-991-258-8	Sequence 8, Appli
21	229.4	3.5	477	10	US-09-981-286A-12	Sequence 12, Appl
22	213.2	3.3	525	10	US-09-981-286A-6	Sequence 6, Appli
23	211.8	3.3	503	10	US-09-981-286A-31	Sequence 31, Appl
24	211.8	3.3	503	10	US-09-981-286A-32	Sequence 32, Appl
25	58.4	0.9	60	12	US-10-123-101-75	Sequence 75, Appl
c 26	46.8	0.7	3673778	12	US-10-312-841-1	Sequence 1, Appli
c 27	46.4	0.7	6239	12	US-10-311-455-1749	Sequence 1749, Ap
c 28	46.4	0.7	6239	12	US-10-240-453-171	Sequence 171, App
c 29	46.2	0.7	15767	12	US-10-311-455-1179	Sequence 1179, Ap
c 30	46.2	0.7	15767	12	US-10-240-485-105	Sequence 105, App
c 31	45	0.7	4057	12	US-10-240-453-261	Sequence 261, App
c 32	45	0.7	4057	14	US-10-239-676-181	Sequence 181, App
c 33	44.2	0.7	7201	12	US-10-311-455-1409	Sequence 1409, Ap
c 34	44	0.7	5930	12	US-10-311-455-1368	Sequence 1368, Ap
35	43.8	0.7	3341	10	US-09-764-868-1433	Sequence 1433, Ap
36	43.8	0.7	3342	10	US-09-764-868-1434	Sequence 1434, Ap
c 37	43.8	0.7	6640	12	US-10-311-455-1630	Sequence 1630, Ap
c 38	43.8	0.7	8238	12	US-10-311-455-1962	Sequence 1962, Ap
c 39	43.6	0.7	398	10	US-09-960-352-11930	Sequence 11930, A
c 40	43.6	0.7	271990	14	US-10-195-144-87	Sequence 87, Appl
c 41	43.4	0.7	7461	12	US-10-311-455-1757	Sequence 1757, Ap
42	43.2	0.7	11416	12	US-10-311-455-91	Sequence 91, Appl
43	43.2	0.7	3673778	12	US-10-312-841-2	Sequence 2, Appli
44	43	0.7	113515	12	US-10-311-455-2147	Sequence 2147, Ap
c 45	43	0.7	3673778	12	US-10-312-841-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-10-023-649-1
; Sequence 1, Application US/10023649
; Publication No. US20030143201A1
; GENERAL INFORMATION:
; APPLICANT: The Minister of National Defence, Government of Canada
; APPLICANT: Nagata, Leslie P
; APPLICANT: Wong, Jonathon P
; TITLE OF INVENTION: No. US20030143201A1el DNA-Based Vaccine Against the Encephalitis
; FILE REFERENCE: NEL-001
; CURRENT APPLICATION NUMBER: US/10/023,649
; CURRENT FILING DATE: 2002-12-27
; PRIOR APPLICATION NUMBER: 60/256,948
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 11484
; TYPE: DNA
; ORGANISM: Western equine encephalomyelitis virus - strain 71V-1658
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (25)..(7428)
; OTHER INFORMATION: 5' UTR <1 .. 24
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7473)..(11183)
; OTHER INFORMATION: encodes nucleocapsid, E3, E2, 6K and E1 proteins
US-10-023-649-1

Query Match 100.0%; Score 6485; DB 12; Length 11484;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CACCGCCAGCAACGACAACCGTAATATGTGAACCAACTGTGCCAGAACGTATACCCAGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5000 CACCGCCAGCAACGACAACCGTAATATGTGAACCAACTGTGCCAGAACGTATACCCAGTC 5059

Qy         61 CGGTGATTTCTAGAGCACCAAGTGCGGAATCACTGCTATCGCTTGGCGGCGTCTCGTTCT 120
```

```
Db       5060 CGGTGATTTCTAGAGCACCAAGTGCGGAATCACTGCTATCGCTTGGCGGCGTCTCGTTCT 5119
Qy        121 CTAGCTCTGCCACACGCTCGTCAACCGCCTGGAGCGACTATGACAGGCGGTTTGTGGTTA 180
Db       5120 CTAGCTCTGCCACACGCTCGTCAACCGCCTGGAGCGACTATGACAGGCGGTTTGTGGTTA 5179
Qy        181 CAGCTGATGTGCATCAAGCGAACACGTCTACGTGGAGCATCCCTAGTGCTCCTGGCTTGG 240
Db       5180 CAGCTGATGTGCATCAAGCGAACACGTCTACGTGGAGCATCCCTAGTGCTCCTGGCTTGG 5239
Qy        241 ACGTCCAGCTGCCTTCTGACGTCACTGATTCCCACTGGAGTATTCCAAGTGCATCAGGCT 300
Db       5240 ACGTCCAGCTGCCTTCTGACGTCACTGATTCCCACTGGAGTATTCCAAGTGCATCAGGCT 5299
Qy        301 TTGAAGTGAGAACACCATCTGTACAGGACCTAACTGCGGAGTGTGCGAAGCCTCGTGGAC 360
Db       5300 TTGAAGTGAGAACACCATCTGTACAGGACCTAACTGCGGAGTGTGCGAAGCCTCGTGGAC 5359
Qy        361 TGGCCGAAATAATGCAAGACTTCAATACTGCTCCTTTCCAGTTTCTTTCGGACTACAGAC 420
Db       5360 TGGCCGAAATAATGCAAGACTTCAATACTGCTCCTTTCCAGTTTCTTTCGGACTACAGAC 5419
Qy        421 CAGTACCGGCACCACGGAGACGCCCCATCCCATCACCTAGATCGACGGCTTCCGCACCTC 480
Db       5420 CAGTACCGGCACCACGGAGACGCCCCATCCCATCACCTAGATCGACGGCTTCCGCACCTC 5479
Qy        481 CAGTTCCAAAGCCACGCAGGACTAAGTACCAACAACCACCAGGAGTCGCTAGAGCGATCT 540
Db       5480 CAGTTCCAAAGCCACGCAGGACTAAGTACCAACAACCACCAGGAGTCGCTAGAGCGATCT 5539
Qy        541 CAGAAGCGGAGTTGGACGAGTACATCCGTCAACACTCCAACTGACGGTATGAAGCGGGAG 600
Db       5540 CAGAAGCGGAGTTGGACGAGTACATCCGTCAACACTCCAACTGACGGTATGAAGCGGGAG 5599
Qy        601 CGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTCAACAGAAATCAGTACGTCAAT 660
Db       5600 CGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTCAACAGAAATCAGTACGTCAAT 5659
Qy        661 GTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATGAGAAGTATTACGCCCCGCGCC 720
Db       5660 GTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATGAGAAGTATTACGCCCCGCGCC 5719
Qy        721 TCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGCAATTATGCGCCTCTGAAGGAA 780
Db       5720 TCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGCAATTATGCGCCTCTGAAGGAA 5779
Qy        781 ATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGAAAGCAATTACAGCGGAGCGAC 840
Db       5780 ATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGAAAGCAATTACAGCGGAGCGAC 5839
Qy        841 TCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGAATCCTGTCGAGTGTTACAGAG 900
Db       5840 TCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGAATCCTGTCGAGTGTTACAGAG 5899
Qy        901 TCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTAACAGGTTTACATCTGCAGAGG 960
Db       5900 TCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTAACAGGTTTACATCTGCAGAGG 5959
Qy        961 TCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATTACCCTACAGTAGCCAGTTATT 1020
Db       5960 TCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATTACCCTACAGTAGCCAGTTATT 6019
Qy       1021 GTATAACAGATGAATACGATGCGTATCTTGACATGGTGGACGGCGCATCGTGCTGTCTAG 1080
Db       6020 GTATAACAGATGAATACGATGCGTATCTTGACATGGTGGACGGCGCATCGTGCTGTCTAG 6079
Qy       1081 ATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACCCAAAGAAGCATAGCTATTTGC 1140
Db       6080 ATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACCCAAAGAAGCATAGCTATTTGC 6139
Qy       1141 AGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGAATACATTACAAAATGTATTGG 1200
Db       6140 AGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGAATACATTACAAAATGTATTGG 6199
Qy       1201 CTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGCGAGAATTACCTGTCTTAGATT 1260
Db       6200 CTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGCGAGAATTACCTGTCTTAGATT 6259
Qy       1261 CGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCATGCAATGATGAGTACTGGGATA 1320
Db       6260 CGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCATGCAATGATGAGTACTGGGATA 6319
Qy       1321 CCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACGTTACGCAATATGTGACAAAGC 1380
Db       6320 CCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACGTTACGCAATATGTGACAAAGC 6379
Qy       1381 TGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTCATAATCTAAAACCGTTGCAGG 1440
Db       6380 TGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTCATAATCTAAAACCGTTGCAGG 6439
Qy       1441 AGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAGATGTCAAAGTTACTCCCGGCA 1500
Db       6440 AGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAGATGTCAAAGTTACTCCCGGCA 6499
Qy       1501 CGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTCAGGCTGCAGATCCCCTTGCTA 1560
Db       6500 CGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTCAGGCTGCAGATCCCCTTGCTA 6559
Qy       1561 CCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTAGACTGAATGCGGTGCTTCTGC 1620
Db       6560 CCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTAGACTGAATGCGGTGCTTCTGC 6619
Qy       1621 CAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATTTTGATGCGATTATTGCTGAAC 1680
Db       6620 CAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATTTTGATGCGATTATTGCTGAAC 6679
Qy       1681 ATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCGCGTCGTTTGATAAAAGCGAAG 1740
Db       6680 ATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCGCGTCGTTTGATAAAAGCGAAG 6739
Qy       1741 ACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGGACTTAGGTGTCGACCAACCGC 1800
Db       6740 ACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGGACTTAGGTGTCGACCAACCGC 6799
Qy       1801 TCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACATCTGTGCACCTACCTACAGGAA 1860
Db       6800 TCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACATCTGTGCACCTACCTACAGGAA 6859
Qy       1861 CGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGTTCTTAACGCTGTTTGTCAACA 1920
Db       6860 CGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGTTCTTAACGCTGTTTGTCAACA 6919
Qy       1921 CACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTGAACGGTTAACCACGTCAGCGT 1980
Db       6920 CACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTGAACGGTTAACCACGTCAGCGT 6979
Qy       1981 GCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTGTCGTCTCCGACACCTTGATGG 2040
Db       6980 GCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTGTCGTCTCCGACACCTTGATGG 7039
Qy       2041 CGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAATTATTGATGCAGTTATTGGTA 2100
Db       7040 CGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAATTATTGATGCAGTTATTGGTA 7099
Qy       2101 TCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGGACCAGATAACAGGCACAGCCT 2160
Db       7100 TCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGGACCAGATAACAGGCACAGCCT 7159
Qy       2161 GCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTGGAAAACCATTGCCAGTCGATG 2220
Db       7160 GCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTGGAAAACCATTGCCAGTCGATG 7219
Qy       2221 ATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATGAAGCAATGCGATGGAACAGAA 2280
Db       7220 ATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATGAAGCAATGCGATGGAACAGAA 7279
```

```
Qy      2281 TTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCAGATACGAGATCATACTGGCAG 2340
Db      7280 TTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCAGATACGAGATCATACTGGCAG 7339

Qy*     2341 GCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCGTTAAGAACTTCAAGAGCATAA 2400
Db      7340 GCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCGTTAAGAACTTCAAGAGCATAA 7399

Qy      2401 GAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGGTGACGTAGTAGACACGCACCT 2460
Db      7400 GAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGGTGACGTAGTAGACACGCACCT 7459

Qy      2461 ACCCACCGGCAGAATGTTTCCATACCCTCAGCTGAACTTTCCACCAGTTTACCCTACAAA 2520
Db      7460 ACCCACCGGCAGAATGTTTCCATACCCTCAGCTGAACTTTCCACCAGTTTACCCTACAAA 7519

Qy      2521 TCCGATGGCTTACCGAGATCCAAACCCTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCC 2580
Db      7520 TCCGATGGCTTACCGAGATCCAAACCCTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCC 7579

Qy      2581 GCTGGCTGCTCAAATCGAAGATCTTAGGAGGTCGATAGTCAACTTGACTTTCAAACAACG 2640
Db      7580 GCTGGCTGCTCAAATCGAAGATCTTAGGAGGTCGATAGTCAACTTGACTTTCAAACAACG 7639

Qy      2641 ATCACCTAATCCGCCGCCAGGTCCACCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAA 2700
Db      7640 ATCACCTAATCCGCCGCCAGGTCCACCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAA 7699

Qy      2701 ACCTACTCAGCCTAAAAAGAAGAAGCAGCAAGCCAAGAGGACGAAACGCAAGCCTAAACC 2760
Db      7700 ACCTACTCAGCCTAAAAAGAAGAAGCAGCAAGCCAAGAGGACGAAACGCAAGCCTAAACC 7759

Qy      2761 AGGGAAACGACAACGTATGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATCATGCT 2820
Db      7760 AGGGAAACGACAACGTATGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATCATGCT 7819

Qy      2821 GAACGGCCAAGTGAATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACTCCA 2880
Db      7820 GAACGGCCAAGTGAATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACTCCA 7879

Qy      2881 CGTTGAAGGAAAAATTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAGCAT 2940
Db      7880 CGTTGAAGGAAAAATTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAGCAT 7939

Qy      2941 GTACGACTTGGAGTACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTACAC 3000
Db      7940 GTACGACTTGGAGTACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTACAC 7999

Qy      3001 CAGCGACAAACCACCGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAATGG 3060
Db      8000 CAGCGACAAACCACCGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAATGG 8059

Qy      3061 GAGATTTACCGTACCGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGA 3120
Db      8060 GAGATTTACCGTACCGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGA 8119

Qy      3121 CAACAGAGGCAGAGTTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGC 3180
Db      8120 CAACAGAGGCAGAGTTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGC 8179

Qy      3181 GCTTTCAGTGGTCACTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTC 3240
Db      8180 GCTTTCAGTGGTCACTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTC 8239

Qy      3241 TGAACCGTGGTCACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGA 3300
Db      8240 TGAACCGTGGTCACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGA 8299

Qy      3301 CAAACCACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAA 3360
Db      8300 CAAACCACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAA 8359

Qy      3361 CGTCGACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAATGTCCATCACGCCG 3420
Db      8360 CGTCGACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAATGTCCATCACGCCG 8419

Qy      3421 GCCCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCCC 3480
Db      8420 GCCCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCCC 8479

Qy      3481 GTATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACGA 3540
Db      8480 GTATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACGA 8539

Qy      3541 ATCTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAGG 3600
Db      8540 ATCTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAGG 8599

Qy      3601 CACTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAAGGA 3660
Db      8600 CACTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAAGGA 8659

Qy      3661 AGACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCACAA 3720
Db      8660 AGACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCACAA 8719

Qy      3721 AGGGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCACGAG 3780
Db      8720 AGGGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCACGAG 8779

Qy      3781 CGGAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTAG 3840
Db      8780 CGGAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTAG 8839

Qy      3841 AGAGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATCA 3900
Db      8840 AGAGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATCA 8899

Qy      3901 CTTGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAA 3960
Db      8900 CTTGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAA 8959

Qy      3961 GTCCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGT 4020
Db      8960 GTCCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGT 9019

Qy      4021 CACCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGAT 4080
Db      9020 CACCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGAT 9079

Qy      4081 GAACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGT 4140
Db      9080 GAACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGT 9139

Qy      4141 CTTCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACAT 4200
Db      9140 CTTCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACAT 9199

Qy      4201 TCCATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCAC 4260
Db      9200 TCCATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCAC 9259

Qy      4261 GAAGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAAC 4320
Db      9260 GAAGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAAC 9319

Qy      4321 GAGAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAG 4380
Db      9320 GAGAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAG 9379

Qy      4381 GAATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAG 4440
Db      9380 GAATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAG 9439

Qy      4441 AGTCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCAT 4500
```

```
Db      9440 AGTCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCAT 9499

Qy      4501 CCACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGC 4560
Db      9500 CCACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGC 9559

Qy      4561 TATCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCT 4620
Db      9560 TATCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCT 9619

Qy      4621 GACGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTG 4680
Db      9620 GACGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTG 9679

Qy      4681 CATTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAA 4740
Db      9680 CATTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAA 9739

Qy      4741 CCAACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCG 4800
Db      9740 CCAACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCG 9799

Qy      4801 CTGCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGA 4860
Db      9800 CTGCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGA 9859

Qy      4861 CGCCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGT 4920
Db      9860 CGCCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGT 9919

Qy      4921 CGAACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAAC 4980
Db      9920 CGAACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAAC 9979

Qy      4981 ACCTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACA 5040
Db      9980 ACCTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACA 10039

Qy      5041 AGTTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCG 5100
Db     10040 AGTTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCG 10099

Qy      5101 CGTTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGA 5160
Db     10100 CGTTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGA 10159

Qy      5161 GAACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGC 5220
Db     10160 GAACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGC 10219

Qy      5221 AGTCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAA 5280
Db     10220 AGTCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAA 10279

Qy      5281 CACCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCT 5340
Db     10280 CACCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCT 10339

Qy      5341 GAAGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCAT 5400
Db     10340 GAAGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCAT 10399

Qy      5401 CAGAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGC 5460
Db     10400 CAGAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGC 10459

Qy      5461 GTTCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACAT 5520
Db     10460 GTTCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACAT 10519

Qy      5521 ACGGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGG 5580
```

```
Db     10520 ACGGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGG 10579

Qy      5581 GTATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATG 5640
Db     10580 GTATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATG 10639

Qy      5641 TAAAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTC 5700
Db     10640 TAAAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTC 10699

Qy      5701 GATTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGT 5760
Db     10700 GATTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGT 10759

Qy      5761 TAGCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACA 5820
Db     10760 TAGCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACA 10819

Qy      5821 GTACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTT 5880
Db     10820 GTACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTT 10879

Qy      5881 GAAGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATC 5940
Db     10880 GAAGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATC 10939

Qy      5941 GAGCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGA 6000
Db     10940 GAGCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGA 10999

Qy      6001 ATGTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCA 6060
Db     11000 ATGTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCA 11059

Qy      6061 GGCGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATC 6120
Db     11060 GGCGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATC 11119

Qy      6121 CCTCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAG 6180
Db     11120 CCTCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAG 11179

Qy      6181 ATGACTGAGCGCGGACACTGACATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGG 6240
Db     11180 ATGACTGAGCGCGGACACTGACATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGG 11239

Qy      6241 TGCATAATGCCACGCGCCGCTTGACACTAAAACTCGATGTATTTCCGAGGAAGCACAGTG 6300
Db     11240 TGCATAATGCCACGCGCCGCTTGACACTAAAACTCGATGTATTTCCGAGGAAGCACAGTG 11299

Qy      6301 CATAATGCTGTGCAGTGTCACATTAATCGTATATCACACTACATATTAACAACACTATAT 6360
Db     11300 CATAATGCTGTGCAGTGTCACATTAATCGTATATCACACTACATATTAACAACACTATAT 11359

Qy      6361 CACTTTTATGAGACTCACTATGGGTCTCTAATATACACTACACATATTTTACTTAAAAAC 6420
Db     11360 CACTTTTATGAGACTCACTATGGGTCTCTAATATACACTACACATATTTTACTTAAAAAC 11419

Qy      6421 ACTATACACACTTTATAAATTCTTTTATAATTTTTCTTTTGTTTTTATTTTGTTTTTAAA 6480
Db     11420 ACTATACACACTTTATAAATTCTTTTATAATTTTTCTTTTGTTTTTATTTTGTTTTTAAA 11479

Qy      6481 ATTTC 6485
Db     11480 ATTTC 11484
```

RESULT 2
US-10-023-649-7
; Sequence 7, Application US/10023649
; Publication No. US20030143201A1
; GENERAL INFORMATION:
; APPLICANT: The Minister of National Defence, Government of Canada

```
;  APPLICANT:  Nagata, Leslie P
;  APPLICANT:  Wong, Jonathon P
;  TITLE OF INVENTION: No. US20030143201A1el DNA-Based Vaccine Against the Encephalitis
;  FILE REFERENCE: NEL-001
;  CURRENT APPLICATION NUMBER: US/10/023,649
;  CURRENT FILING DATE:  2002-12-27
;  PRIOR APPLICATION NUMBER: 60/256,948
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
;   LENGTH: 4395
;   TYPE: DNA
;   ORGANISM: Western equine encephalomyelitis virus - STRAIN 71v-1658
;   FEATURE:
;   NAME/KEY: CMV promoter
;   LOCATION: (1)..(1260)
;   OTHER INFORMATION: Pvax vector sequence: 1-196; CMV promoter: 1-115; CMV putative tr
;   OTHER INFORMATION: anscriptional start site: 125; T7 promoter: 48-167; pVAX multiclo
;   OTHER INFORMATION: ning region: 168-196; polyprotein (C-E3-E2-6K-E1): 214-4065; pcDW
;   OTHER INFORMATION: -HX45 nontranslated region: 4066-4348; pcDW-HX45 vector sequence:
;   OTHER INFORMATION: 4349-4385; pVAX vector sequence: 4386
US-10-023-649-7

  Query Match           63.8%;  Score 4137.2;  DB 12;  Length 4395;
  Best Local Similarity  99.9%;  Pred. No. 0;
  Matches 4139;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2324 CGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCGTTAA 2383
             |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 CGAGCTCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCGTTAA 266

Qy       2384 GAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGGTGAC 2443
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 GAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGGTGAC 326

Qy       2444 GTAGTAGACACGCACCTACCCACCGGCAGAATGTTTCCATACCCTCAGCTGAACTTTCCA 2503
             |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        327 GTAGTAGACACGCACCTACCCACCGCCAGAATGTTTCCATACCCTCAGCTGAACTTTCCA 386

Qy       2504 CCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACCCTCCTAGGCGCCGCTGG 2563
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 CCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACCCTCCTAGGCGCCGCTGG 446

Qy       2564 AGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTAGGAGGTCGATAGTCAAC 2623
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        447 AGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTAGGAGGTCGATAGTCAAC 506

Qy       2624 TTGACTTTCAAACAACGATCACCTAATCCGCCGCCAGGTCCACCGCCAAAGAAGAAGAAG 2683
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 TTGACTTTCAAACAACGATCACCTAATCCGCCGCCAGGTCCACCGCCAAAGAAGAAGAAG 566

Qy       2684 AGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGAAGAAGCAGCAAGCCAAGAGGACG 2743
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        567 AGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGAAGAAGCAGCAAGCCAAGAGGACG 626

Qy       2744 AAACGCAAGCCTAAACCAGGGAAACGACAACGTATGTGTATGAAGTTGGAGTCGGACAAG 2803
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        627 AAACGCAAGCCTAAACCAGGGAAACGACAACGTATGTGTATGAAGTTGGAGTCGGACAAG 686

Qy       2804 ACATTTCCGATCATGCTGAACGGCCAAGTGAATGGATATGCCTGCGTTGTCGGAGGAAGG 2863
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        687 ACATTTCCGATCATGCTGAACGGCCAAGTGAATGGATATGCCTGCGTTGTCGGAGGAAGG 746

Qy       2864 CTGATGAAACCACTCCACGTTGAAGGAAAAATTGATAATGAGCAATTAGCGGCCGTGAAA 2923
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        747 CTGATGAAACCACTCCACGTTGAAGGAAAAATTGATAATGAGCAATTAGCGGCCGTGAAA 806

Qy       2924 TTGAAGAAGGCTAGCATGTACGACTTGGAGTACGGCGACGTTCCCCAGAACATGAAATCA 2983
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        807 TTGAAGAAGGCTAGCATGTACGACTTGGAGTACGGCGACGTTCCCCAGAACATGAAATCA 866

Qy       2984 GACACGCTGCAGTACACCAGCGACAAACCACCGGGCTTCTACAACTGGCACCACGGCGCA 3043
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        867 GACACGCTGCAGTACACCAGCGACAAACCACCGGGCTTCTACAACTGGCACCACGGCGCA 926

Qy       3044 GTCCAGTATGAGAATGGGAGATTTACCGTACCGAGAGGAGTGGGCGGGAAAGGCGACAGC 3103
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        927 GTCCAGTATGAGAATGGGAGATTTACCGTACCGAGAGGAGTGGGCGGGAAAGGCGACAGC 986

Qy       3104 GGAAGACCGATCCTGGACAACAGAGGCAGAGTTGTGGCTATTGTTCTAGGAGGTGCAAAT 3163
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        987 GGAAGACCGATCCTGGACAACAGAGGCAGAGTTGTGGCTATTGTTCTAGGAGGTGCAAAT 1046

Qy       3164 GAGGGCACGCGTACGGCGCTTTCAGTGGTCACTTGGAACCAGAAAGGGGTGACCATTAGG 3223
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1047 GAGGGCACGCGTACGGCGCTTTCAGTGGTCACTTGGAACCAGAAAGGGGTGACCATTAGG 1106

Qy       3224 GATACCCCCGAAGGTTCTGAACCGTGGTCACTAGTTACAGCGCTATGCGTGCTTTCGAAT 3283
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1107 GATACCCCCGAAGGTTCTGAACCGTGGTCACTAGTTACAGCGCTATGCGTGCTTTCGAAT 1166

Qy       3284 GTCACGTTCCCATGCGACAAACCACCCGTGTGCTATTCACTGACGCCAGAACGAACACTC 3343
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1167 GTCACGTTCCCATGCGACAAACCACCCGTGTGCTATTCACTGACGCCAGAACGAACACTC 1226

Qy       3344 GACGTGCTCGAAGAGAACGTCGACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTG 3403
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1227 GACGTGCTCGAAGAGAACGTCGACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTG 1286

Qy       3404 AAATGTCCATCACGCCGGCCCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCC 3463
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1287 AAATGTCCATCACGCCGGCCCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCC 1346

Qy       3464 TACCTGGGGTTCTGCCCGTATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATT 3523
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1347 TACCTGGGGTTCTGCCCGTATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATT 1406

Qy       3524 GAGAACGTGTGGGACGAATCTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTC 3583
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1407 GAGAACGTGTGGGACGAATCTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTC 1466

Qy       3584 GGCTACAATCAGGCAGGCACTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCAC 3643
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1467 GGCTACAATCAGGCAGGCACTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCAC 1526

Qy       3644 GACCATGACATCAAGGAAGACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGC 3703
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1527 GACCATGACATCAAGGAAGACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGC 1586

Qy       3704 CGTCGTCTTGGCCACAAAGGGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTA 3763
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1587 CGTCGTCTTGGCCACAAAGGGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTA 1646

Qy       3764 ACCGTCAGTATCACGAGCGGAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGG 3823
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1647 ACCGTCAGTATCACGAGCGGAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGG 1706

Qy       3824 AGGAAGTTTGTCGGTAGAGAGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAG 3883
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1707 AGGAAGTTTGTCGGTAGAGAGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAG 1766

Qy       3884 TGCCACGTTTACGATCACTTGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCA 3943
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1767 TGCCACGTTTACGATCACTTGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCA 1826

Qy       3944 GGCCCACACGCGTATAAGTCCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCA 4003
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1827 GGCCCACACGCGTATAAGTCCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCA 1886

Qy       4004 CCTTCTGGCAAGAACGTCACCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTG 4063
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1887 CCTTCTGGCAAGAACGTCACCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTG 1946

Qy       4064 AGCACGCGAACGAAGATGAACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGC 4123
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

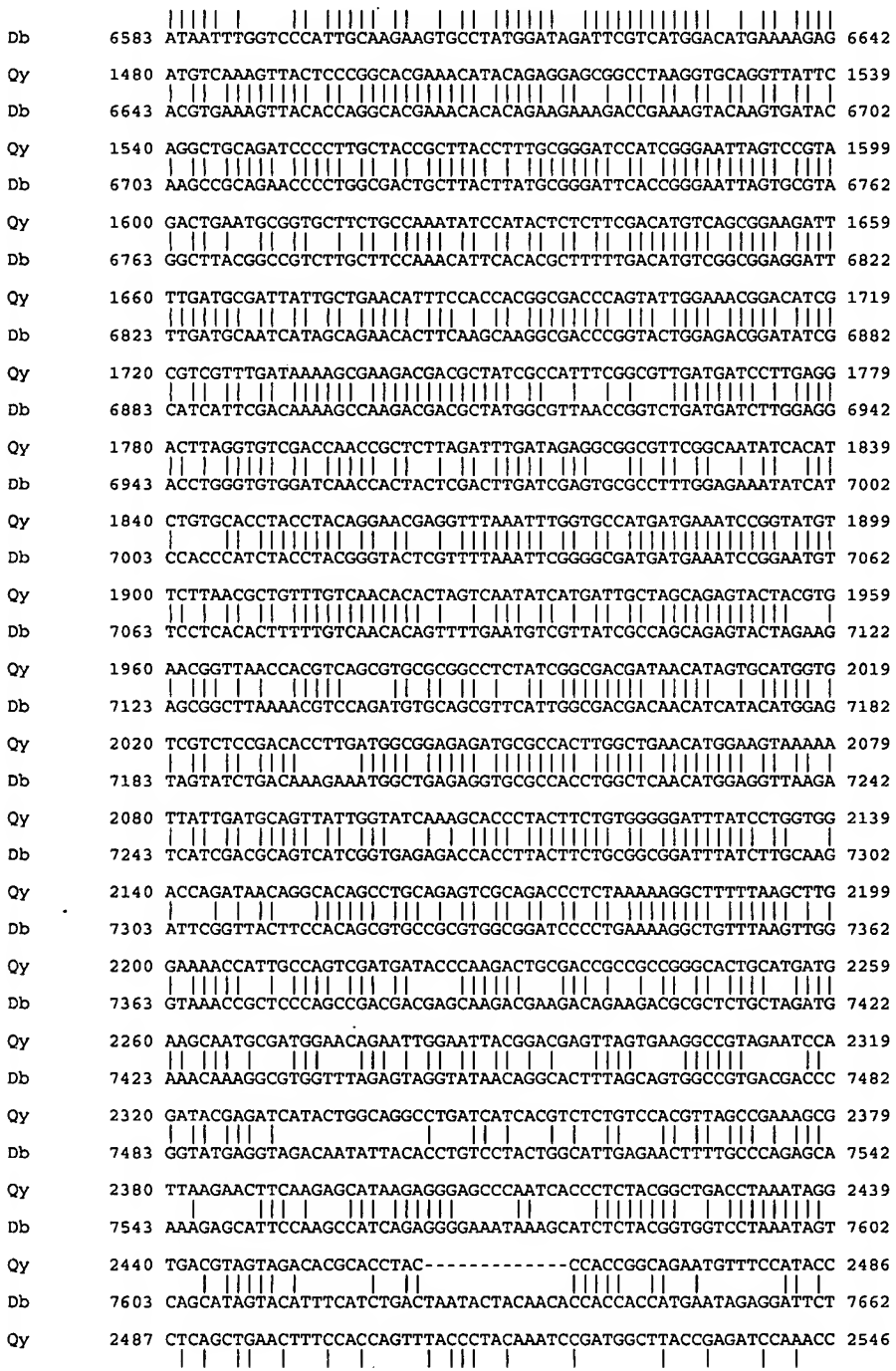
```
Db      1947 AGCACGCGAACGAAGATGAACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGC 2006
Qy      4124 GACCAAACGAAATGGGTCTTCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTG 4183
Db      2007 GACCAAACGAAATGGGTCTTCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTG 2066
Qy      4184 CAAGGTAAATTGCACATTCCATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCT 4243
Db      2067 CAAGGTAAATTGCACATTCCATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCT 2126
Qy      4244 CACACGCCTACAGTCACGAAGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGA 4303
Db      2127 CACACGCCTACAGTCACGAAGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGA 2186
Qy      4304 CCAACATTGCTGACAACGAGAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATT 4363
Db      2187 CCAACATTGCTGACAACGAGAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATT 2246
Qy      4364 ACAGGGTCTACATCCAGGAATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGT 4423
Db      2247 ACAGGGTCTACATCCAGGAATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGT 2306
Qy      4424 AACCATGAACCAGTCAGAGTCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGG 4483
Db      2307 AACCATGAACCAGTCAGAGTCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGG 2366
Qy      4484 CCGCATGAGATCATCATCCACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTG 4543
Db      2367 CCGCATGAGATCATCATCCACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTG 2426
Qy      4544 TGTGGTGTCGCTCTTGCTATCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAA 4603
Db      2427 TGTGGTGTCGCTCTTGCTATCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAA 2486
Qy      4604 GCAAGAAGAGACTGCCTGACGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCA 4663
Db      2487 GCAAGAAGAGACTGCCTGACGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCA 2546
Qy      4664 TTAGCGGTTTTGTGCTGCATTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAAC 4723
Db      2547 TTAGCGGTTTTGTGCTGCATTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAAC 2606
Qy      4724 CATCTGTGGTTTAACAACCAACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCG 4783
Db      2607 CATCTGTGGTTTAACAACCAACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCG 2666
Qy      4784 CTTGTTATTCTGTTCCGCTGCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTC 4843
Db      2667 CTTGTTATTCTGTTCCGCTGCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTC 2726
Qy      4844 TGCCTGGGGAAGGTAGACGCCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATC 4903
Db      2727 TGCCTGGGGAAGGTAGACGCCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATC 2786
Qy      4904 CCGTATAAGGCGTTGGTCGAACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTC 4963
Db      2787 CCGTATAAGGCGTTGGTCGAACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTC 2846
Qy      4964 GTCTCATCGGAATTAACACCTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACA 5023
Db      2847 GTCTCATCGGAATTAACACCTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACA 2906
Qy      5024 GTCATTCCTTCACCACAAGTTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAG 5083
Db      2907 GTCATTCCTTCACCACAAGTTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAG 2966
Qy      5084 GCGGATTACACATGCCGCGTTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAA 5143
Db      2967 GCGGATTACACATGCCGCGTTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAA 3026
Qy      5144 TGCTTCTGTGACAGTGAGAACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGAC 5203
Db      3027 TGCTTCTGTGACAGTGAGAACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGAC 3086

Qy      5204 TGCACTATAGATCACGCAGTCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTG 5263
Db      3087 TGCACTATAGATCACGCAGTCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTG 3146
Qy      5264 CGTATAGTATACGGCAACACCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCA 5323
Db      3147 CGTATAGTATACGGCAACACCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCA 3206
Qy      5324 GGTTCCTCACGGGACCTGAAGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTT 5383
Db      3207 GGTTCCTCACGGGACCTGAAGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTT 3266
Qy      5384 GACCATAAGGTCGTCATCAGAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGA 5443
Db      3267 GACCATAAGGTCGTCATCAGAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGA 3326
Qy      5444 GCTATGAAACCAGGAGCGTTCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATA 5503
Db      3327 GCTATGAAACCAGGAGCGTTCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATA 3386
Qy      5504 GTAGCCCGCACTGACATACGGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTAC 5563
Db      3387 GTAGCCCGCACTGACATACGGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTAC 3446
Qy      5564 ACCCAAGCAGTATCAGGGTATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAA 5623
Db      3447 ACCCAAGCAGTATCAGGGTATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAA 3506
Qy      5624 ACAGCACCATTTGGATGTAAAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTAC 5683
Db      3507 ACAGCACCATTTGGATGTAAAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTAC 3566
Qy      5684 GGGCACATCCCTATCTCGATTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCA 5743
Db      3567 GGGCACATCCCTATCTCGATTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCA 3626
Qy      5744 CCAACAATTTTAGAAGTTAGCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGT 5803
Db      3627 CCAACAATTTTAGAAGTTAGCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGT 3686
Qy      5804 GGTTCTCTAACATTACAGTACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCAC 5863
Db      3687 GGTTCTCTAACATTACAGTACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCAC 3746
Qy      5864 TCCACGACAGCTGTTTTGAAGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACA 5923
Db      3747 TCCACGACAGCTGTTTTGAAGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACA 3806
Qy      5924 CTACATTTTAGCACATCGAGCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAG 5983
Db      3807 CTACATTTTAGCACATCGAGCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAG 3866
Qy      5984 TCCACCTGCAATGCTGAATGTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAA 6043
Db      3867 TCCACCTGCAATGCTGAATGTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAA 3926
Qy      6044 GTCGACCAAGAATTCCAGGCGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTG 6103
Db      3927 GTCGACCAAGAATTCCAGGCGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTG 3986
Qy      6104 TTTGGGGGAGCATCATCCCTCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATG 6163
Db      3987 TTTGGGGGAGCATCATCCCTCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATG 4046
Qy      6164 CTTATAAACACACGTAGATGACTGAGCGCGGACACTGACATAGCGGTAAAACTCGATGTA 6223
Db      4047 CTTATAAACACACGTAGATGACTGAGCGCGGACACTGACATAGCGGTAAAACTCGATGTA 4106
Qy      6224 CTTCCGAGGAAGCGTGGTGCATAATGCCACGCGCCGCTTGACACTAAAACTCGATGTATT 6283
Db      4107 CTTCCGAGGAAGCGTGGTGCATAATGCCACGCGCCGCTTGACACTAAAACTCGATGTATT 4166
```

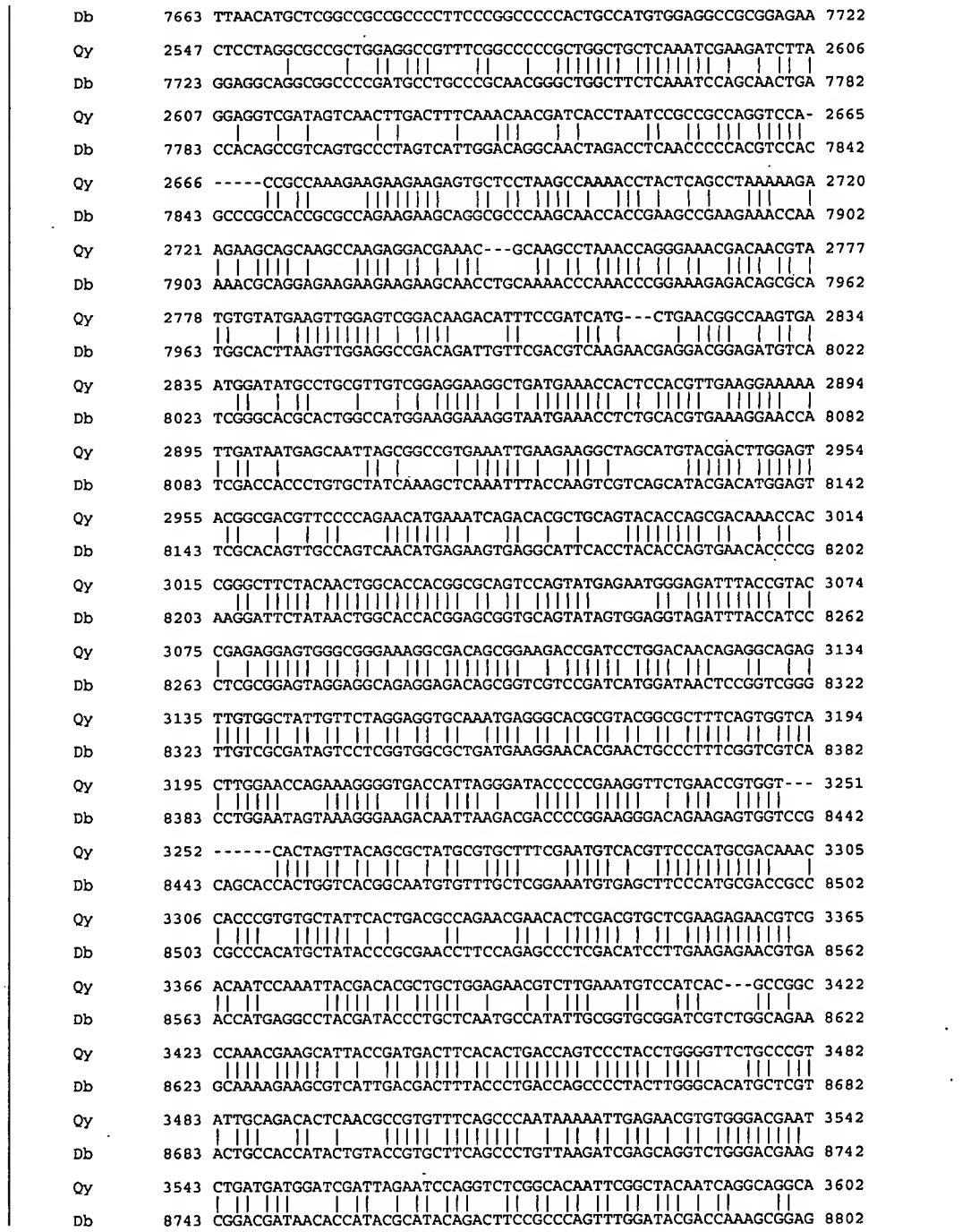
```
Qy       6284 TCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGTCACATTAATCGTATATCACACTACA 6343
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4167 TCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGTCACATTAATCGTATATCACACTACA 4226

Qy       6344 TATTAACAACACTATATCACTTTTATGAGACTCACTATGGGTCTCTAATATACACTACAC 6403
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4227 TATTAACAACACTATATCACTTTTATGAGACTCACTATGGGTCTCTAATATACACTACAC 4286

Qy       6404 ATATTTTACTTAAAAACACTATACACACTTTATAAATTCTTTTATAATTTTTCTTTTGTT 6463
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db       4287 ATATTTTACTTAAAAACACTATACACACTTTATAAATTCTTTTATAATTTTTCTTTTGCT 4346

Qy       6464 TT 6465
              ||
Db       4347 TT 4348


RESULT 3
US-10-023-649-5
; Sequence 5, Application US/10023649
; Publication No. US20030143201A1
; GENERAL INFORMATION:
;  APPLICANT: The Minister of National Defence, Government of Canada
;  APPLICANT:  Nagata, Leslie P
;  APPLICANT:  Wong, Jonathon P
;  TITLE OF INVENTION: No. US20030143201A1el DNA-Based Vaccine Against the Encephalitis
;  FILE REFERENCE: NEL-001
;  CURRENT APPLICATION NUMBER: US/10/023,649
;  CURRENT FILING DATE:  2002-12-27
;  PRIOR APPLICATION NUMBER: 60/256,948
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
;   LENGTH: 4150
;   TYPE: DNA
;   ORGANISM: Western equine encephalomyelitis virus - strain 71V-1658
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (159)..(3869)
;   OTHER INFORMATION: vector sequence 1-9; 5' SacI primer 9-20; 3' end of NS4 gene 16-1
;   OTHER INFORMATION: 14; intragenic region 115-158; polyprotein (C-E3-E2-6K-E1) 159-38
;   OTHER INFORMATION: 56; pcDW-XH7 nontranslated region 3857-4150
US-10-023-649-5

  Query Match             63.7%;  Score 4129.8;  DB 12;  Length 4150;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 4134;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       2325 GAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCGTTAAG 2384
             ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         10 GAGCTCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCGTTAAG 69

Qy       2385 AACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGGTGACG 2444
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 AACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGGTGACG 129

Qy       2445 TAGTAGACACGCACCTACCCACCGGCAGAATGTTTCCATACCCTCAGCTGAACTTTCCAC 2504
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 TAGTAGACACGCACCTACCCACCGGCAGAATGTTTCCATACCCTCAGCTGAACTTTCCAC 189

Qy       2505 CAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACCCTCCTAGGCGCCGCTGGA 2564
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 CAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACCCTCCTAGGCGCCGCTGGA 249

Qy       2565 GGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTAGGAGGTCGATAGTCAACT 2624
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 GGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTAGGAGGTCGATAGTCAACT 309

Qy       2625 TGACTTTCAAACAACGATCACCTAATCCGCCGCCAGGTCCACCGCCAAAGAAGAAGAAGA 2684
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 TGACTTTCAAACAACGATCACCTAATCCGCCGCCAGGTCCACCGCCAAAGAAGAAGAAGA 369

Qy       2685 GTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGAAGAAGCAGCAAGCCAAGAGGACGA 2744
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         370 GTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGAAGAAGCAGCAAGCCAAGAGGACGA 429

Qy       2745 AACGCAAGCCTAAACCAGGGAAACGACAACGTATGTGTATGAAGTTGGAGTCGGACAAGA 2804
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         430 AACGCAAGCCTAAACCAGGGAAACGACAACGTATGTGTATGAAGTTGGAGTCGGACAAGA 489

Qy       2805 CATTTCCGATCATGCTGAACGGCCAAGTGAATGGATATGCCTGCGTTGTCGGAGGAAGGC 2864
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         490 CATTTCCGATCATGCTGAACGGCCAAGTGAATGGATATGCCTGCGTTGTCGGAGGAAGGC 549

Qy       2865 TGATGAAACCACTCCACGTTGAAGGAAAAATTGATAATGAGCAATTAGCGGCCGTGAAAT 2924
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         550 TGATGAAACCACTCCACGTTGAAGGAAAAATTGATAATGAGCAATTAGCGGCCGTGAAAT 609

Qy       2925 TGAAGAAGGCTAGCATGTACGACTTGGAGTACGGCGACGTTCCCCAGAACATGAAATCAG 2984
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         610 TGAAGAAGGCTAGCATGTACGACTTGGAGTACGGCGACGTTCCCCAGAACATGAAATCAG 669

Qy       2985 ACACGCTGCAGTACACCAGCGACAAACCACCGGGCTTCTACAACTGGCACCACGGCGCAG 3044
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         670 ACACGCTGCAGTACACCAGCGACAAACCACCGGGCTTCTACAACTGGCACCACGGCGCAG 729

Qy       3045 TCCAGTATGAGAATGGGAGATTTACCGTACCGAGAGGAGTGGGCGGGAAAGGCGACAGCG 3104
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         730 TCCAGTATGAGAATGGGAGATTTACCGTACCGAGAGGAGTGGGCGGGAAAGGCGACAGCG 789

Qy       3105 GAAGACCGATCCTGGACAACAGAGGCAGAGTTGTGGCTATTGTTCTAGGAGGTGCAAATG 3164
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         790 GAAGACCGATCCTGGACAACAGAGGCAGAGTTGTGGCTATTGTTCTAGGAGGTGCAAATG 849

Qy       3165 AGGGCACGCGTACGGCGCTTTCAGTGGTCACTTGGAACCAGAAAGGGGTGACCATTAGGG 3224
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         850 AGGGCACGCGTACGGCGCTTTCAGTGGTCACTTGGAACCAGAAAGGGGTGACCATTAGGG 909

Qy       3225 ATACCCCCGAAGGTTCTGAACCGTGGTCACTAGTTACAGCGCTATGCGTGCTTTCGAATG 3284
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         910 ATACCCCCGAAGGTTCTGAACCGTGGTCACTAGTTACAGCGCTATGCGTGCTTTCGAATG 969

Qy       3285 TCACGTTCCCATGCGACAAACCACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCG 3344
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         970 TCACGTTCCCATGCGACAAACCACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCG 1029

Qy       3345 ACGTGCTCGAAGAGAACGTCGACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGA 3404
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1030 ACGTGCTCGAAGAGAACGTCGACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGA 1089

Qy       3405 AATGTCCATCACGCCGGCCCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCT 3464
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1090 AATGTCCATCACGCCGGCCCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCT 1149

Qy       3465 ACCTGGGGTTCTGCCCGTATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTG 3524
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1150 ACCTGGGGTTCTGCCCGTATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTG 1209

Qy       3525 AGAACGTGTGGGACGAATCTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCG 3584
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1210 AGAACGTGTGGGACGAATCTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCG 1269

Qy       3585 GCTACAATCAGGCAGGCACTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACG 3644
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1270 GCTACAATCAGGCAGGCACTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACG 1329

Qy       3645 ACCATGACATCAAGGAAGACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCC 3704
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1330 ACCATGACATCAAGGAAGACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCC 1389

Qy       3705 GTCGTCTTGGCCACAAAGGGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAA 3764
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1390 GTCGTCTTGGCCACAAAGGGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAA 1449

Qy       3765 CCGTCAGTATCACGAGCGGAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGA 3824
```

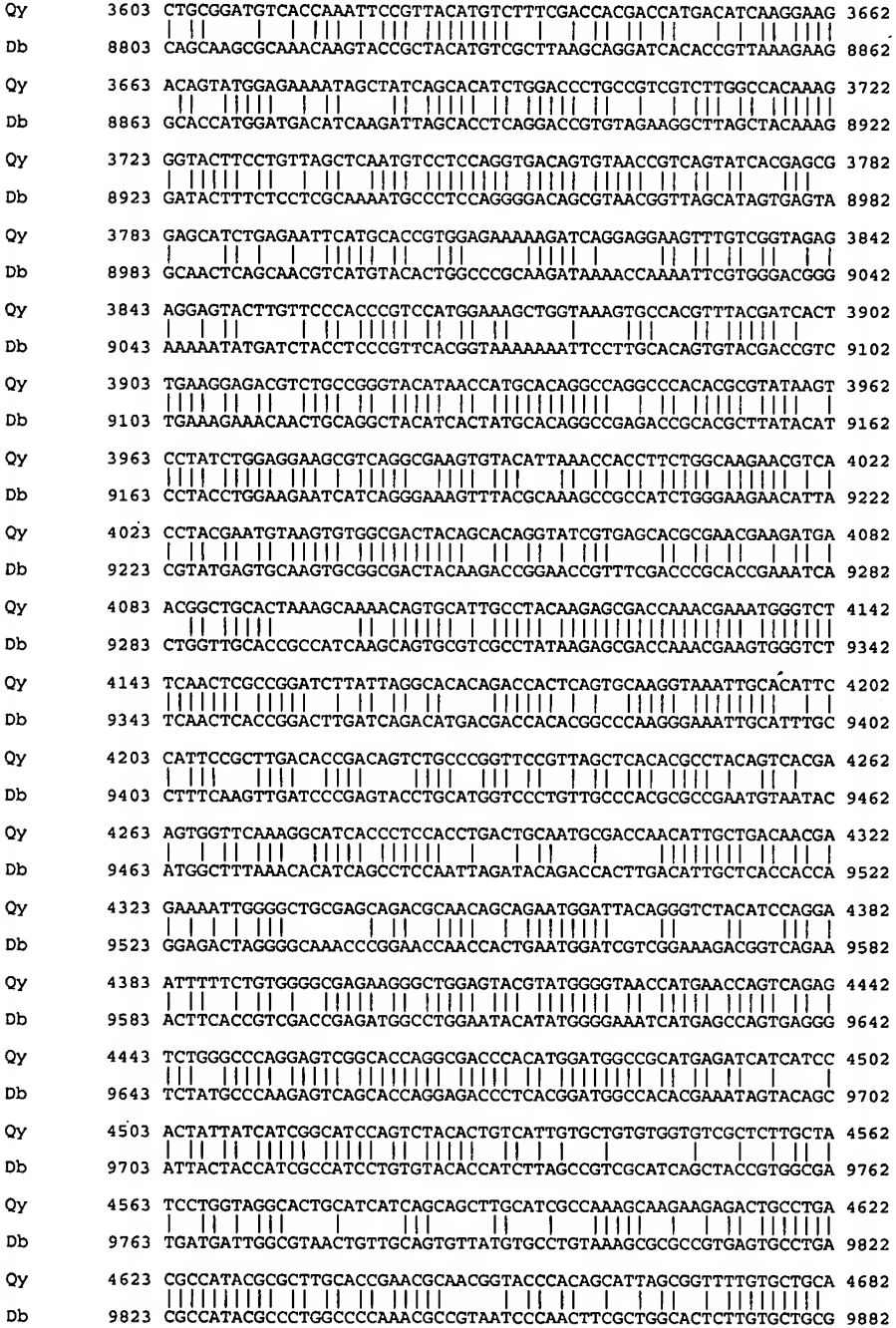
```
Db      1450 CCGTCAGTATCACGAGCGGAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGA 1509
Qy      3825 GGAAGTTTGTCGGTAGAGAGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGT 3884
Db      1510 GGAAGTTTGTCGGTAGAGAGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGT 1569
Qy      3885 GCCACGTTTACGATCACTTGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAG 3944
Db      1570 GCCACGTTTACGATCACTTGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAG 1629
Qy      3945 GCCCACACGCGTATAAGTCCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCAC 4004
Db      1630 GCCCACACGCGTATAAGTCCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCAC 1689
Qy      4005 CTTCTGGCAAGAACGTCACCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGA 4064
Db      1690 CTTCTGGCAAGAACGTCACCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGA 1749
Qy      4065 GCACGCGAACGAAGATGAACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCG 4124
Db      1750 GCACGCGAACGAAGATGAACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCG 1809
Qy      4125 ACCAAACGAAATGGGTCTTCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGC 4184
Db      1810 ACCAAACGAAATGGGTCTTCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGC 1869
Qy      4185 AAGGTAAATTGCACATTCCATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTC 4244
Db      1870 AAGGTAAATTGCACATTCCATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTC 1929
Qy      4245 ACACGCCTACAGTCACGAAGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGAC 4304
Db      1930 ACACGCCTACAGTCACGAAGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGAC 1989
Qy      4305 CAACATTGCTGACAACGAGAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTA 4364
Db      1990 CAACATTGCTGACAACGAGAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTA 2049
Qy      4365 CAGGGTCTACATCCAGGAATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTA 4424
Db      2050 CAGGGTCTACATCCAGGAATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTA 2109
Qy      4425 ACCATGAACCAGTCAGAGTCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGC 4484
Db      2110 ACCATGAACCAGTCAGAGTCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGC 2169
Qy      4485 CGCATGAGATCATCATCCACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGT 4544
Db      2170 CGCATGAGATCATCATCCACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGT 2229
Qy      4545 GTGGTGTCGCTCTTGCTATCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAG 4604
Db      2230 GTGGTGTCGCTCTTGCTATCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAG 2289
Qy      4605 CAAGAAGAGACTGCCTGACGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCAT 4664
Db      2290 CAAGAAGAGACTGCCTGACGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCAT 2349
Qy      4665 TAGCGGTTTTGTGCTGCATTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACC 4724
Db      2350 TAGCGGTTTTGTGCTGCATTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACC 2409
Qy      4725 ATCTGTGGTTTAACAACCAACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGC 4784
Db      2410 ATCTGTGGTTTAACAACCAACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGC 2469
Qy      4785 TTGTTATTCTGTTCCGCTGCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCT 4844
Db      2470 TTGTTATTCTGTTCCGCTGCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCT 2529
Qy      4845 GCCTGGGGAAGGTAGACGCCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCC 4904
Db      2530 GCCTGGGGAAGGTAGACGCCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCC 2589
Qy      4905 CGTATAAGGCGTTGGTCGAACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCG 4964
Db      2590 CGTATAAGGCGTTGGTCGAACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCG 2649
Qy      4965 TCTCATCGGAATTAACACCTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAG 5024
Db      2650 TCTCATCGGAATTAACACCTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAG 2709
Qy      5025 TCATTCCTTCACCACAAGTTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGG 5084
Db      2710 TCATTCCTTCACCACAAGTTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGG 2769
Qy      5085 CGGATTACACATGCCGCGTTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAAT 5144
Db      2770 CGGATTACACATGCCGCGTTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAAT 2829
Qy      5145 GCTTCTGTGACAGTGAGAACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACT 5204
Db      2830 GCTTCTGTGACAGTGAGAACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACT 2889
Qy      5205 GCACTATAGATCACGCAGTCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGC 5264
Db      2890 GCACTATAGATCACGCAGTCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGC 2949
Qy      5265 GTATAGTATACGGCAACACCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAG 5324
Db      2950 GTATAGTATACGGCAACACCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAG 3009
Qy      5325 GTTCCTCACGGGACCTGAAGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTG 5384
Db      3010 GTTCCTCACGGGACCTGAAGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTG 3069
Qy      5385 ACCATAAGGTCGTCATCAGAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAG 5444
Db      3070 ACCATAAGGTCGTCATCAGAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAG 3129
Qy      5445 CTATGAAACCAGGAGCGTTCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAG 5504
Db      3130 CTATGAAACCAGGAGCGTTCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAG 3189
Qy      5505 TAGCCCGCACTGACATACGGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACA 5564
Db      3190 TAGCCCGCACTGACATACGGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACA 3249
Qy      5565 CCCAAGCAGTATCAGGGTATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAA 5624
Db      3250 CCCAAGCAGTATCAGGGTATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAA 3309
Qy      5625 CAGCACCATTTGGATGTAAAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACG 5684
Db      3310 CAGCACCATTTGGATGTAAAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACG 3369
Qy      5685 GGCACATCCCTATCTCGATTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCAC 5744
Db      3370 GGCACATCCCTATCTCGATTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCAC 3429
Qy      5745 CAACAATTTTAGAAGTTAGCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTG 5804
Db      3430 CAACAATTTTAGAAGTTAGCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTG 3489
Qy      5805 GTTCTCTAACATTACAGTACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACT 5864
Db      3490 GTTCTCTAACATTACAGTACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACT 3549
Qy      5865 CCACGACAGCTGTTTTGAAGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACAC 5924
Db      3550 CCACGACAGCTGTTTTGAAGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACAC 3609
Qy      5925 TACATTTTAGCACATCGAGCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGT 5984
Db      3610 TACATTTTAGCACATCGAGCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGT 3669
```

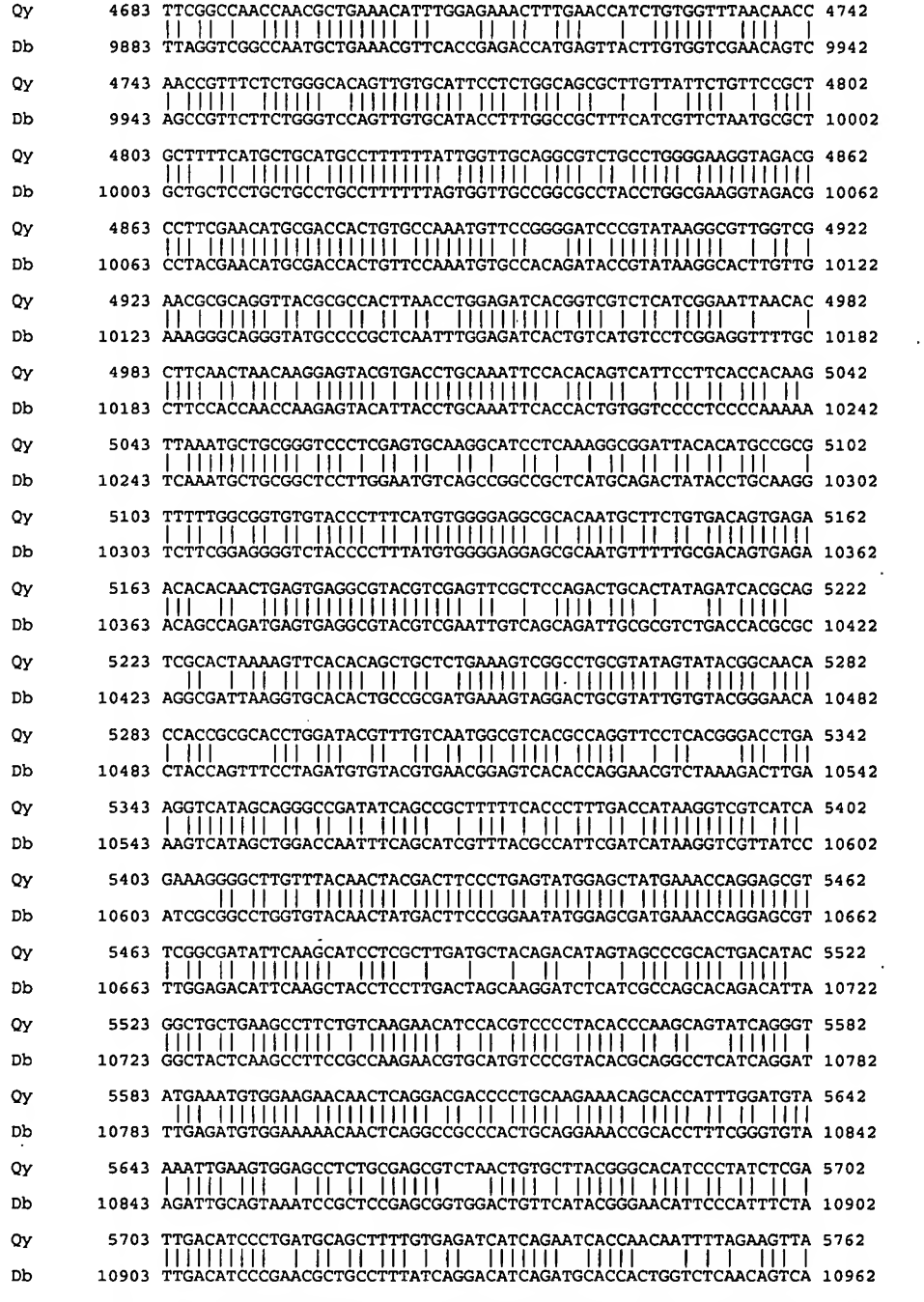
```
Qy       5985 CCACCTGCAATGCTGAATGTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAG 6044
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3670 CCACCTGCAATGCTGAATGTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAG 3729

Qy       6045 TCGACCAAGAATTCCAGGCGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGT 6104
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3730 TCGACCAAGAATTCCAGGCGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGT 3789

Qy       6105 TTGGGGGAGCATCATCCCTCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGC 6164
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3790 TTGGGGGAGCATCATCCCTCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGC 3849

Qy       6165 TTATAAACACACGTAGATGACTGAGCGCGGACACTGACATAGCGGTAAAACTCGATGTAC 6224
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3850 TTATAAACACACGTAGATGACTGAGCGCGGACACTGACATAGCGGTAAAACTCGATGTAC 3909

Qy       6225 TTCCGAGGAAGCGTGGTGCATAATGCCACGCGCCGCTTGACACTAAAACTCGATGTATTT 6284
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3910 TTCCGAGGAAGCGTGGTGCATAATGCCACGCGCCGCTTGACACTAAAACTCGATGTATTT 3969

Qy       6285 CCGAGGAAGCACAGTGCATAATGCTGTGCAGTGTCACATTAATCGTATATCACACTACAT 6344
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3970 CCGAGGAAGCACAGTGCATAATGCTGTGCAGTGTCACATTAATCGTATATCACACTACAT 4029

Qy       6345 ATTAACAACACTATATCACTTTTATGAGACTCACTATGGGTCTCTAATATACACTACACA 6404
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4030 ATTAACAACACTATATCACTTTTATGAGACTCACTATGGGTCTCTAATATACACTACACA 4089

Qy       6405 TATTTTACTTAAAAACACTATACACACTTTATAAATTCTTTTATAATTTTTCTTTTGTTT 6464
             |||||||||||||||||||||||||||||||||| | |||||||  |||| | ||
Db       4090 TATTTTACTTAAAAACACTATACACACTTTATAAATTCTCTCATAATTTCACTTTAGGTT 4149

Qy       6465 T 6465
             |
Db       4150 T 4150
```

```
RESULT 4
US-09-507-362-103
; Sequence 103, Application US/09507362
; Publication No. US20030096397A1
;   GENERAL INFORMATION:
;        APPLICANT: Dubensky Jr., Thomas W.
;                   Polo, John M.
;                   Belli, Barbara A.
;                   Schlesinger, Sondra
;                   Dryga, Sergey A.
;                   Frolov, Ilya
;        TITLE OF INVENTION: RECOMBINANT ALPHAVIRUS-BASED VECTORS
;                            WITH REDUCED INHIBITION OF CELLULAR MACRO-MOLECULAR
;                            SYNTHESIS
;        NUMBER OF SEQUENCES: 125
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Seed Intellectual Property Law Group PLLC
;             STREET: 701 Fifth Avenue, Suite 6300
;             CITY: Seattle
;             STATE: Washington
;             COUNTRY: USA
;             ZIP: 98104-7092
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/507,362
;             FILING DATE: 18-Feb-2000
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McMasters, David D.
;             REGISTRATION NUMBER: 33,963
;             REFERENCE/DOCKET NUMBER: 930049.457D6 /1196.011
;             TELECOMMUNICATION INFORMATION:
;                  TELEPHONE: (206) 622-4900
;                  TELEFAX: (206) 682-6031
;   INFORMATION FOR SEQ ID NO: 103:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 11740 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 103:
US-09-507-362-103

  Query Match             36.7%;  Score 2380.4;  DB 11;  Length 11740;
  Best Local Similarity   64.6%;  Pred. No. 0;
  Matches 3693;  Conservative    0;  Mismatches 1986;  Indels   41;  Gaps    8;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
              ||||||     |  |  || | ||| || || ||  |||| |||||||  || ||| | |
Db       5746 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5805

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
              || |||| || || |  ||     | ||   |||||| |  ||||| ||    ||||  |
Db       5806 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTCCTGG 5865

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
              | |     | | ||||||  |||||         | |||| ||     |  |      |
Db       5866 AAAGAATTCATGCCCCGGTGCTCGACACGTCGAAAGAGGAACAACTCAAACTCAGGTACC 5925

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
              | |     || | ||||  || | ||| |||||| || || || ||||||||||||  ||
Db       5926 AGATGATGCCCACCGAAGCCAACAAAAGTAGGTACCAGTCTCGTAAAGTAGAAAATCAGA 5985

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
              |||| || ||  | ||||||||  | || ||| |  |  ||| || |  ||
Db       5986 AAGCCATAACCACTGAGCGACTACTGTCAGGACTACGACTGTATAACTCTGCCACAG--- 6042

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
              |||      || || || |   |||  |||||  |||| | |||||    |  |||
Db       6043 ATCAGCCAGAATGCTATAAGATCACCTATCCGAAACCATTGTACTCCAGTAGCGTACCGG 6102

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
                |  |   |  | || || |||| ||        || |||     | | || ||||| |
Db       6103 CGAACTACTCCGATCCACAGTTCGCTGTAGCTGTCTGTAACAACTATCTGCATGAGAACT 6162

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
              | || ||||||||   ||||   || || || || |||||||| ||  | || ||||| |
Db       6163 ATCCGACAGTAGCATCTTATCAGATTACTGACGAGTACGATGCTTACTTGGATATGGTAG 6222

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
              ||||  ||     ||| || ||||| || || || || || ||||| || ||||| ||||
Db       6223 ACGGGACAGTCGCCTGCCTGGATACTGCAACCTTCTGCCCCGCTAAGCTTAGAAGTTACC 6282

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
              | || || |||   |||      ||  | ||  |   || || |||||  | || ||||
Db       6283 CGAAAAAACATGAGTATAGAGCCCCGAATATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6342

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
              | ||  |||||||||  |   ||| || |||||||| |||||||||| || || ||||
Db       6343 ACACGCTACAAAATGTGCTCATTGCCGCAACTAAAAGAAATTGCAACGTCACGCAGATGC 6402

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
              | ||| | ||      | || || ||| |||| |||| || || ||     |||| ||||
Db       6403 GTGAACTGCCAACACTGGACTCAGCGACATTCAATGTCGAATGCTTTCGAAAATATGCAT 6462

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
              | ||||| ||||| ||||    ||     || || ||| || | || || |||  |
Db       6463 GTAATGACGAGTATTGGGAGGAGTTCGCTCGGAAGCCAATTAGGATTACCACTGAGTTTG 6522

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
              | ||   ||||||  | |  |||||||| || || || || ||| | ||||| || ||
Db       6523 TCACCGCATATGTAGCTAGACTGAAAGGCCCTAAGGCCGCCGCACTATTTGCAAAGACGT 6582

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
```

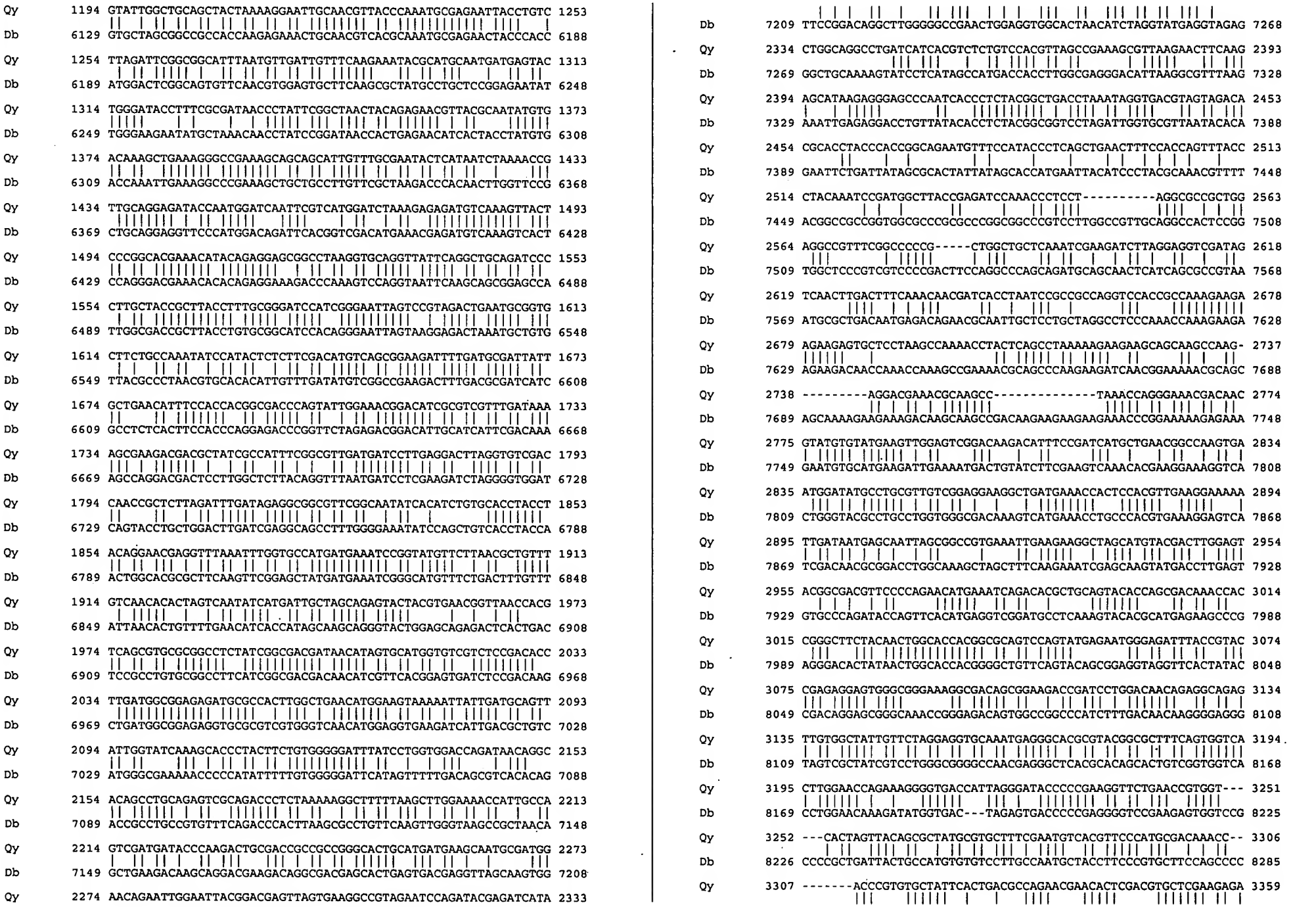








```
Db      10963 AATGTGAAGTCAGTGAGTGCACTTATTCAGCAGACTTCGGCGGGATGGCCACCCTGCAGT 11022

Qy       5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882

Db      11023 ATGTATCCGACCGCGAAGGTCAATGCCCCGTACATTCGCATTCGAGCACAGCAACTCTCC 11082

Qy       5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942

Db      11083 AAGAGTCGACAGTACATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTTAGCACCGCGA 11142

Qy       5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002

Db      11143 GTCCACAGGCGAACTTTATCGTATCGCTGTGTGGGAAGAAGACAACATGCAATGCAGAAT 11202

Qy       6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062

Db      11203 GTAAACCACCAGCTGACCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTTCAAG 11262

Qy       6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122

Db      11263 CCGCCATCTCAAAAACATCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11322

Qy       6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182

Db      11323 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11382

Qy       6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241

Db      11383 GACCGCTACGCCCCAATGATCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11442

Qy       6242 GCATAATGCCACGCGCCGCT 6261

Db      11443 GCATAATGCATCAGGCTGGT 11462


RESULT 5
US-09-901-106-1
; Sequence 1, Application US/09901106
; Patent No. US20020151067A1
;   GENERAL INFORMATION:
;        APPLICANT: Garoff, Henrik
;                   Liljestrom, Peter
;        TITLE OF INVENTION: DNA Expression Systems Based on
;                            Alphaviruses
;        NUMBER OF SEQUENCES: 27
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Birch, Stewart, Kolasch & Birch
;             STREET: P.O. Box 747
;             CITY: Falls Church
;             STATE: Virginia
;             COUNTRY: USA
;             ZIP: 22040-0747
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/901,106
;             FILING DATE: 10-Jul-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/07/920,281C
;             FILING DATE: 13-AUG-1992
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Murphy Jr., Gerald M.
;             REGISTRATION NUMBER: 28,977
;             REFERENCE/DOCKET NUMBER: 828-103P
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 703-241-1300
;             TELEFAX: 703-241-2848
;             TELEX: 248345
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 11517 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: RNA (genomic)
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        ORIGINAL SOURCE:
;             ORGANISM: Semliki Forest Virus
;        FEATURE:
;             NAME/KEY:  -
;             LOCATION:  1..11517
;             OTHER INFORMATION:  /label= genome
;             /note= "Semliki Forest Virus complete nucleotide
;             sequence, presented as a cloned DNA sequence; see
;             Figure 5."
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  87..7379
;             OTHER INFORMATION:  /product= "SFV polyprotein"
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  7421..11179
;             OTHER INFORMATION:  /product= "SFV polyprotein"
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-901-106-1

  Query Match             22.2%;  Score 1442.6;  DB 10;  Length 11517;
  Best Local Similarity   55.5%;  Pred. No. 0;
  Matches 3130;  Conservative    0;  Mismatches 2409;  Indels  102;  Gaps   14;

Qy        594 GCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTCAACAGAAATCAGTA 653

Db       5529 GCGGGTGCATATATTTTCTCCTCGGACACTGGCAGCGGACATTTACAACAAAAATCCGTT 5588

Qy        654 CGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATGAGAAGTATTACGCC 713

Db       5589 AGGCAGCACAATCTCCAGTGCGCACAACTGGATGCGGTCCAGGAGGAGAAAATGTACCCG 5648

Qy        714 CCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGCAATTATGCGCCTCT 773

Db       5649 CCAAAATTGGATACTGAGAGGGAGAAGCTGTTGCTGCTGAAAATGCAGATGCACCCATCG 5708

Qy        774 GAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGAAAGCAATTACAGCG 833

Db       5709 GAGGCTAATAAGAGTCGATACCAGTCTCGCAAAGTGGAGAACATGAAAGCCACGGTGGTG 5768

Qy        834 GAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGAATCCTGTCGAGTGT 893

Db       5769 GACAGGCTCACATCGGGGGCCAGATTGTACACGGGAGCGGACGTAGGCCGCATACCAACA 5828

Qy        894 TACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTAACAGGTTTACATCT 953

Db       5829 TACGCGGTTCGGTACCCCCGCCCCGTGTACTCCCCTACCGTGATCGAAAGATTCTCAAGC 5888

Qy        954 GCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATTACCCTACAGTAGCC 1013

Db       5889 CCCGATGTAGCAATCGCAGCGTGCAACGAATACCTATCCAGAAATTACCCAACAGTGGCG 5948

Qy       1014 AGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGGACGGCGCATCGTGC 1073

Db       5949 TCGTACCAGATAACAGATGAATACGACGCATACTTGGACATGGTTGACGGGTCGGATAGT 6008

Qy       1074 TGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACCCAAAGAAGCATAGC 1133

Db       6009 TGCTTGGACAGAGCGACATTCTGCCCGGCGAAGCTCCGGTGCTACCCGAAACATCATGCG 6068

Qy       1134 TATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGAATACATTACAAAAT 1193

Db       6069 TACCACCAGCCGACTGTACGCAGTGCCGTCCCGTCACCCTTTCAGAACACACTACAGAAC 6128
```



```
Db       8286 CGTGTGTACCTTGCTGCTATGAAAACAACGCAGAGGCCACACTACGGATGCTCGAGGATA 8345
Qy       3360 ACGTCGACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAATGTCCA---TCAC 3416
Db       8346 ACGTGGATAGGCCAGGGTACTACGACCTCCTTCAGGCAGCCTTGACGTGCCGAAACGGAA 8405
Qy       3417 GCCGGCCCAAACGAAGCATTACCGATGACTTCACACTG---------ACCAGTCCCTACC 3467
Db       8406 CAAGACACCGGCGCAGCGTGTCGCAACACTTCAACGTGTATAAGGCTACACGCCCTTACA 8465
Qy       3468 TGGGGTTCTGCCCGTATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGA 3527
Db       8466 TCGCGTACTGCGCCGACTGCGGAGCAGGGCACTCGTGTCATAGCCCCGTAGCAATTGAAG 8525
Qy       3528 ACGTGTGGGACGAATCTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCT 3587
Db       8526 CGGTCAGGTCCGAAGCTACCGACGGGATGCTGAAGATTCAGTTCTCGGCACAAATTGGCA 8585
Qy       3588 ACAATCAGGCAGGCACTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACC 3647
Db       8586 TAGATAAGAGTGACAATCATGACTACACGAAGATAAGGTA---------CGCAGACGGGC 8636
Qy       3648 ATGACATCAAGGAAGACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTC 3707
Db       8637 ACGCCATTGAGAATGCCGTCCGGTCATCTTTGAAGGTAGCCACCTCCGGAGACTGTTTCG 8696
Qy       3708 GTCTTGGCCACAAAGGGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCG 3767
Db       8697 TCCATGGCACAATGGGACATTTCATACTGGCAAAGTGCCCACCGGGTGAATTCCTGCAGG 8756
Qy       3768 TCAGTATCACG---AGCGGAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGA 3824
Db       8757 TCTCGATCCAGGACACCAGAAACGCGGTCCGTGCCTGCAGAATACAATATCATCATGACC 8816
Qy       3825 GGAAGTTTGTCGGTAGAGAGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGT 3884
Db       8817 CTCAACCGGTGGGTAGAGAAAAATTTACAATTAGACCACACTATGGAAAAGAGATCCCTT 8876
Qy       3885 GCCACGTTTACGATCACTTGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAG 3944
Db       8877 GCACCACTTATCAACAGACCACAGCGAAGACCGTGGAGGAAATCGACATGCATATGCCGC 8936
Qy       3945 GCCCACACGCGTATAAGTCCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCAC 4004
Db       8937 CAGATACGCCGGACAGGACGTTGCTATCACAGCAATCTGGCAATGTAAAGATCA---CAG 8993
Qy       4005 CTTCTGGCAAGAACGTCACCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGA 4064
Db       8994 TCGGAGGAAAGAAGGTGAAATACAACTGCACCTGTGGAACCGGAAACGTTGGCACTACTA 9053
Qy       4065 GCACGCGAACGAAGATGAACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCG 4124
Db       9054 ATTCGGACATGACGATCAACACGTG---TCTAATAGAGCAGTGCCACGTCTCAGTGACGG 9110
Qy       4125 ACCAAACGAAATGGGTCTTCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGC 4184
Db       9111 ACCATAAGAAATGGCAGTTCAACTCACCTTTCGTCCCGAGAGCCGACGAACCGGCTAGAA 9170
Qy       4185 AAGGTAAATTGCACATTCCATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTC 4244
Db       9171 AAGGCAAAGTCCATATCCCATTCCCGTTGGACAACATCACATGCAGAGTTCCAATGGCGC 9230
Qy       4245 ACACGCCTACAGTCACGAAGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGAC 4304
Db       9231 GCGAACCAACCGTCATCCACGGCAAAAGAGAAGTGACACTGCACCTTCACCCAGATCATC 9290
Qy       4305 CAACATTGCTGACAACGAGAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTA 4364
Db       9291 CCACGCTCTTTTCCTACCGCACACTGGGTGAGGACCCGCAGTATCACGAGGAATGGGTGA 9350
Qy       4365 CAGGGTCTACATCCAGGAATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTA 4424
Db       9351 CAGCGGCGGTGGAACGGACCATACCCGTACCAGTGGACGGGATGGAGTACCACTGGGGAA 9410

Qy       4425 ACCATGAACCAGTCAGAGTCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGC 4484
Db       9411 ACAACGACCCAGTGAGGCTTTGGTCTCAACTCACCACTGAAGGGAAACCGCACGGCTGGC 9470
Qy       4485 CGCATGAGATCATCATCCACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGT 4544
Db       9471 CGCATCAGATCGTACAGTACTACTATGGGCTTTACCCGGCCGCTACAGTATCCGCGGTCG 9530
Qy       4545 GTGGTGTCGCTCTTGCTATCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAG 4604
Db       9531 TCGGGATGAGCTTACTGGCGTTGATATCGATCTTCGCGTCGTGCTACATGCTGGTTGCGG 9590
Qy       4605 CAAGAAGAGACTGCCTGACGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCAT 4664
Db       9591 CCCGCAGTAAGTGCTTGACCCCTTATGCTTTAACACCAGGAGCTGCAGTTCCGTGGACGC 9650
Qy       4665 TAGCGGTTTTGTGCTGCATTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACC 4724
Db       9651 TGGGGATACTCTGCTGCGCCCCGCGGGCGCACGCAGCTAGTGTGGCAGAGACTATGGCCT 9710
Qy       4725 ATCTGTGGTTTAACAACCAACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGC 4784
Db       9711 ACTTGTGGGACCAAAACCAAGCGTTGTTCTGGTTGGAGTTTGCGGCCCCTGTTGCCTGCA 9770
Qy       4785 TTGTTATTCTGTTCCGCTGC---------------TTTTCATGCTGCATGCCTTTTTTAT 4829
Db       9771 TCCTCATCATCACGTATTGCCTCAGAAACGTGCTGTGTTGCTGTAAGAGCCTTTCTTTTT 9830
Qy       4830 TGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACGCCTTCGAACATGCGACCACTGTGCCAA 4889
Db       9831 TAGTGCTACTGAGCCTCGGGGCAACCGCCAGAGCTTACGAACATTCGACAGTAATGCCGA 9890
Qy       4890 ATGTTCCGGGGATCCCGTATAAGGCGTTGGTCGAACGCGCAGGTTACGCGCCACTTAACC 4949
Db       9891 ACGTGGTGGGGTTCCCGTATAAGGCTCACATTGAAAGGCCAGGATATAGCCCCCTCACTT 9950
Qy       4950 TGGAGATCACGGTCGTCTCATCGGAATTAACACCTTCAACTAACAAGGAGTACGTGACCT 5009
Db       9951 TGCAGATGCAGGTTGTTGAAACCAGCCTCGAACCAACCCTTAATTTGGAATACATAACCT 10010
Qy       5010 GCAAATTCCACACAGTCATTCCTTCACCACAAGTTAAATGCTGCGGGTCCCTCGAGTGCA 5069
Db       10011 GTGAGTACAAGACGGTCGTCCCGTCGCCGTACGTGAAGTGCTGCGGCGCCTCAGAGTGCT 10070
Qy       5070 AGGCATCCTCAAAGGCGGATTACACATGCCGCGTTTTTGGCGGTGTGTACCCTTTCATGT 5129
Db       10071 CCACTAAAGAGAAGCCTGACTACCAATGCAAGGTTTACACAGGCGTGTACCCGTTCATGT 10130
Qy       5130 GGGGAGGCGCACAATGCTTCTGTGACAGTGAGAACACACAACTGAGTGAGGCGTACGTCG 5189
Db       10131 GGGGAGGGGCATATTGCTTCTGCGACTCAGAAAACACGCAACTCAGCGAGGCGTACGTCG 10190
Qy       5190 AGTTCGCTCCAGACTGCACTATAGATCACGCAGTCGCACTAAAAGTTCACACAGCTGCTC 5249
Db       10191 ATCGATCGGACGTATGCAGGCATGATCACGCATCTGCTTACAAAGCCCATACAGCATCGC 10250
Qy       5250 TGAAAGTCGGCCTGCGTATAGTATACGGCAACACCACCGCGCACCTGGATACGTTTGTCA 5309
Db       10251 TGAAGGCCAAAGTGAGGGTTATGTACGGCAACGTAAACCAGACTGTGGATGTTTACGTGA 10310
Qy       5310 ATGGCGTCACGCCAGGTTCCTCACGGGACCTGAAGGTCATAGCAGGGCCGATATCAGCCG 5369
Db       10311 ACGGAGACCATGCCGTCACGATAGGGGGTACTCAGTTCATATTCGGGCCGCTGTCATCGG 10370
Qy       5370 CTTTTTCACCCTTTGACCATAAGGTCGTCATCAGAAAGGGGCTTGTTTACAACTACGACT 5429
Db       10371 CCTGGACCCCGTTCGACAACAAGATAGTCGTGTACAAAGACGAAGTGTTCAATCAGGACT 10430
Qy       5430 TCCCTGAGTATGGAGCTATGAAACCAGGAGCGTTCGGCGATATTCAAGCATCCTCGCTTG 5489
Db       10431 TCCCGCCGTACGGATCTGGGCAACCAGGGCGCTTCGGCGACATCCAAAGCAGAACAGTGG 10490
```

```
Qy       5490 ATGCTACAGACATAGTAGCCCGCACTGACATACGGCTGCTGAAGCCTTCTGTCAAGAACA 5549
Db      10491 AGAGTAACGACCTGTACGCGAACACGGCACTGAAGCTGGCACGCCCTTCACCCGGCATGG 10550

Qy       5550 TCCACGTCCCCTACACCCAAGCAGTATCAGGGTATGAAATGTGGAAGAACAACTCAGGAC 5609
Db      10551 TCCATGTACCGTACACACAGACACCTTCAGGGTTCAAATATTGGCTAAAGGAAAAAGGGA 10610

Qy       5610 GACCCCTGCAAGAAACAGCACCATTTGGATGTAAAATTGAAGTGGAGCCTCTGCGAGCGT 5669
Db      10611 CAGCCCTAAATACGAAGGCTCCTTTTGGCTGCCAAATCAAAACGAACCCTGTCAGGGCCA 10670

Qy       5670 CTAACTGTGCTTACGGGCACATCCCTATCTCGATTGACATCCCTGATGCAGCTTTTGTGA 5729
Db      10671 TGAACTGCGCCGTGGGAAACATCCCTGTCTCCATGAATTTGCCTGACAGCGCCTTTACCC 10730

Qy       5730 GATCATCAGAATCACCAACAATTTTAGAAGTTAGCTGCACAGTAGCAGACTGCATTTATT 5789
Db      10731 GCATTGTCGAGGCGCCGACCATCATTGACCTGACTTGCACAGTGGCTACCTGTACGCACT 10790

Qy       5790 CTGCAGACTTTGGTGGTTCTCTAACATTACAGTACAAAGCTGACAGGGAGGGACATTGTC 5849
Db      10791 CCTCGGATTTCGGCGGCGTCTTGACACTGACGTACAAGACCAACAAGAACGGGGACTGCT 10850

Qy       5850 CAGTTCACTCCCACTCCACGACAGCTGTTTTGAAGGAAGCGACCACACATGTGACTGCCG 5909
Db      10851 CTGTACACTCGCACTCTAACGTAGCTACTCTACAGGAGGCCACAGCAAAAGTGAAGACAG 10910

Qy       5910 TAGGCAGCATAACACTACATTTTAGCACATCGAGCCCACAAGCAAATTTTATAGTTTCGC 5969
Db      10911 CAGGTAAGGTGACCTTACACTTCTCCACGGCAAGCGCATCACCTTCTTTTGTGGTGTCGC 10970

Qy       5970 TATGCGGCAAGAAGTCCACCTGCAATGCTGAATGTAAACCACCGGCCGACCACATAATTG 6029
Db      10971 TATGCAGTGCTAGGGCCACCTGTTCAGCGTCGTGTGAGCCCCCGAAAGACCACATAGTCC 11030

Qy       6030 GAGAACCACATAAAGTCGACCAAGAATTCCAGGCGGCAGTTTCCAAAACATCTTGGAACT 6089
Db      11031 CATATGCGGCTAGCCACAGTAACGTAGTGTTTCCAGACATGTCGGGCACCGCACTATCAT 11090

Qy       6090 GGCTGCTTGCACTGTTTGGGGGAGCATCATCCCTCATTGTTGTAGGACTTATAGTGTTGG 6149
Db      11091 GGGTGCAGAAAATCTCGGGTGGTCTGGGGGCCTTCGCAATCGGCGCTATCCTGGTGCTGG 11150

Qy       6150 T 6150
Db      11151 T 11151
```

```
RESULT 6
US-09-991-258-1
; Sequence 1, Application US/09991258
; Patent No. US20020141975A1
; GENERAL INFORMATION:
;  APPLICANT: Olmsted, Robert
;  APPLICANT:  Keith, Paula
;  APPLICANT:  Dryga, Sergey
;  APPLICANT:  Caley, Ian
;  APPLICANT:  Maughan, Maureen
;  APPLICANT:  Johnston, Robert
;  APPLICANT:  Davis, Nancy
;  APPLICANT:  Swanstrom, Ronald
;  TITLE OF INVENTION: ALPHAVIRUS VECTORS AND VIROSOMES WITH MODIFIED HIV GENES FOR USE
;  TITLE OF INVENTION:  VACCINES
;  FILE REFERENCE: 01113.0001U3
;  CURRENT APPLICATION NUMBER: US/09/991,258
;  CURRENT FILING DATE:  2001-11-16
;  PRIOR APPLICATION NUMBER: 09/902,537
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: 60/216,995
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 4.0
```

```
; SEQ ID NO 1
;   LENGTH: 12523
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence; No. US20020141975A1e =
;   OTHER INFORMATION: synthetic construct
US-09-991-258-1

  Query Match           14.3%;  Score 926.2;  DB 10;  Length 12523;
  Best Local Similarity   68.1%;  Pred. No. 2.2e-265;
  Matches 1320;  Conservative    0;  Mismatches  613;  Indels    6;  Gaps    2;

Qy        514 AACCACCAGGAGTCGCTAGAGCGATCTCAGAAGCGGAGTTGGACGAGTACATCCGTCAAC 573
Db       5617 ACCCGCCAGGCGTAAATAGGGTGATTACAAGAGAGGAGTTTGAGGCGTTCGTAGCACAAC 5676

Qy        574 ACTCCAACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTC 633
Db       5677 A---ACAATGACGGTTTGATGCGGGTGCATACATCTTTTCCTCCGACACCGGTCAAGGGC 5733

Qy        634 ACCTTCAACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCG 693
Db       5734 ATTTACAACAAAAATCAGTAAGGCAAACGGTGCTATCCGAAGTGGTGTTGGAGAGGACCG 5793

Qy        694 TCCATGAGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGA 753
Db       5794 AATTGGAGATTTCGTATGCCCCGCGCCTCGACCAAGAAAAAGAAGAATTACTACGCAAGA 5853

Qy        754 AACTGCAATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAA 813
Db       5854 AATTACAGTTAAATCCCACACCTGCTAACAGAAGCAGATACCAGTCCAGGAAGGTGGAGA 5913

Qy        814 ATATGAAAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAG 873
Db       5914 ACATGAAAGCCATAACAGCTAGACGTATTCTGCAAGGCCTAGGGCATTATTTGAAGGCAG 5973

Qy        874 AAGTGAATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGG 933
Db       5974 AAGGAAA---AGTGGAGTGCTACCGAACCCTGCATCCTGTTCCTTTGTATTCATCTAGTG 6030

Qy        934 TAATTAACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAG 993
Db       6031 TGAACCGTGCCTTTTCAAGCCCCAAGGTCGCAGTGGAAGCCTGTAACGCCATGTTGAAAG 6090

Qy        994 AGAATTACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACA 1053
Db       6091 AGAACTTTCCGACTGTGGCTTCTTACTGTATTATTCCAGAGTACGATGCCTATTTGGACA 6150

Qy       1054 TGGTGGACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAA 1113
Db       6151 TGGTTGACGGAGCTTCATGCTGCTTAGACACTGCCAGTTTTTGCCCTGCAAAGCTGCGCA 6210

Qy       1114 GCTACCCAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTA 1173
Db       6211 GCTTTCCAAAGAAACACTCCTATTTGGAACCCACAATACGATCGGCAGTGCCTTCAGCGA 6270

Qy       1174 TACAGAATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCC 1233
Db       6271 TCCAGAACACGCTCCAGAACGTCCTGGCAGCTGCCACAAAAAGAAATTGCAATGTCACGC 6330

Qy       1234 AAATGCGAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAAT 1293
Db       6331 AAATGAGAGAATTGCCCGTATTGGATTCGGCGGCCTTTAATGTGGAATGCTTCAAGAAAT 6390

Qy       1294 ACGCATGCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAG 1353
Db       6391 ATGCGTGTAATAATGAATATTGGGAAACGTTTAAAGAAAACCCCATCAGGCTTACTGAAG 6450

Qy       1354 AGAACGTTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGA 1413
Db       6451 AAAACGTGGTAAATTACATTACCAAATTAAAAGGACCAAAAGCTGCTGCTCTTTTTGCGA 6510
```

```
Qy       1414 ATACTCATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAA 1473
Db       6511 AGACACATAATTTGAATATGTTGCAGGACATACCAATGGACAGGTTTGTAATGGACTTAA 6570

Qy       1474 AGAGAGATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGG 1533
Db       6571 AGAGAGACGTGAAAGTGACTCCAGGAACAAAACATACTGAAGAACGGCCCAAGGTACAGG 6630

Qy       1534 TTATTCAGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAG 1593
Db       6631 TGATCCAGGCTGCCGATCCGCTAGCAACAGCGTATCTGTGCGGAATCCACCGAGAGCTGG 6690

Qy       1594 TCCGTAGACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGG 1653
Db       6691 TTAGGAGATTAAATGCGGTCCTGCTTCCGAACATTCATACACTGTTTGATATGTCGGCTG 6750

Qy       1654 AAGATTTTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGG 1713
Db       6751 AAGACTTTGACGCTATTATAGCCGAGCACTTCCAGCCTGGGGATTGTGTTCTGGAAACTG 6810

Qy       1714 ACATCGCGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCC 1773
Db       6811 ACATCGCGTCGTTTGATAAAAGTGAGGACGACGCCATGGCTCTGACCGCGTTAATGATTC 6870

Qy       1774 TTGAGGACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATA 1833
Db       6871 TGGAAGACTTAGGTGTGGACGCAGAGCTGTTGACGCTGATTGAGGCGGCTTTCGGCGAAA 6930

Qy       1834 TCACATCTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCG 1893
Db       6931 TTTCATCAATACATTTGCCCACTAAAACTAAATTTAAATTCGGAGCCATGATGAAATCTG 6990

Qy       1894 GTATGTTCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTAC 1953
Db       6991 GAATGTTCCTCACACTGTTTGTGAACACAGTCATTAACATTGTAATCGCAAGCAGAGTGT 7050

Qy       1954 TACGTGAACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGC 2013
Db       7051 TGAGAGAACGGCTAACCGGATCACCATGTGCAGCATTCATTGGAGATGACAATATCGTGA 7110

Qy       2014 ATGGTGTCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAG 2073
Db       7111 AAGGAGTCAAATCGGACAAATTAATGGCAGACAGGTGCGCCACCTGGTTGAATATGGAAG 7170

Qy       2074 TAAAAATTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCC 2133
Db       7171 TCAAGATTATAGATGCTGTGGTGGGCGAGAAAGCGCCCTATTTCTGTGGAGGGTTTATTT 7230

Qy       2134 TGGTGGACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTA 2193
Db       7231 TGTGTGACTCCGTGACCGGCACAGCGTGCCGTGTGGCAGACCCCCTAAAAAGGCTGTTTA 7290

Qy       2194 AGCTTGGAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGC 2253
Db       7291 AGCTTGGCAAACCTCTGGCAGCAGACGATGAACATGATGATGACAGGAGAAGGGCATTGC 7350

Qy       2254 ATGATGAAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAG 2313
Db       7351 ATGAAGAGTCAACACGCTGGAACCGAGTGGGTATTCTTTCAGAGCTGTGCAAGGCAGTAG 7410

Qy       2314 AATCCAGATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCG 2373
Db       7411 AATCAAGGTATGAAACCGTAGGAACTTCCATCATAGTTATGGCCATGACTACTCTAGCTA 7470

Qy       2374 AAAGCGTTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTA 2433
Db       7471 GCAGTGTTAAATCATTCAGCTACCTGAGAGGGGCCCCTATAACTCTCTACGGCTAACCTG 7530

Qy       2434 AATAGGTGACGTAGTAGAC 2452
Db       7531 AATGGACTACGACATAGTC 7549
```

```
RESULT 7
US-09-991-258-14
; Sequence 14, Application US/09991258
; Patent No. US20020141975A1
; GENERAL INFORMATION:
;  APPLICANT: Olmsted, Robert
;  APPLICANT:  Keith, Paula
;  APPLICANT:  Dryga, Sergey
;  APPLICANT:  Caley, Ian
;  APPLICANT:  Maughan, Maureen
;  APPLICANT:  Johnston, Robert
;  APPLICANT:  Davis, Nancy
;  APPLICANT:  Swanstrom, Ronald
;  TITLE OF INVENTION: ALPHAVIRUS VECTORS AND VIROSOMES WITH MODIFIED HIV GENES FOR USE I
;  TITLE OF INVENTION:  VACCINES
;  FILE REFERENCE: 01113.0001U3
;  CURRENT APPLICATION NUMBER: US/09/991,258
;  CURRENT FILING DATE:  2001-11-16
;  PRIOR APPLICATION NUMBER: 09/902,537
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: 60/216,995
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
;   LENGTH: 12379
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence; No. US20020141975A1e =
;   OTHER INFORMATION: synthetic construct
US-09-991-258-14

  Query Match             14.3%;  Score 924.6;  DB 10;  Length 12379;
  Best Local Similarity   68.0%;  Pred. No. 6.5e-265;
  Matches 1319;  Conservative    0;  Mismatches  614;  Indels    6;  Gaps    2;

Qy        514 AACCACCAGGAGTCGCTAGAGCGATCTCAGAAGCGGAGTTGGACGAGTACATCCGTCAAC 573
Db       5617 ACCCGCCAGGCGTAAATAGGGTGATTACAAGAGAGGAGTTTGAGGCGTTCGTAGCACAAC 5676

Qy        574 ACTCCAACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTC 633
Db       5677 A---ACAATGACGGTTTGATGCGGGTGCATACATCTTTTCCTCCGACACCGGTCAAGGGC 5733

Qy        634 ACCTTCAACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCG 693
Db       5734 ATTTACAACAAAAATCAGTAAGGCAAACGGTGCTATCCGAAGTGGTGTTGGAGAGGACCG 5793

Qy        694 TCCATGAGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGA 753
Db       5794 AATTGGAGATTTCGTATGCCCCGCGCCTCGACCAAGAAAAAGAAGAATTACTACGCAAGA 5853

Qy        754 AACTGCAATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAA 813
Db       5854 AATTACAGTTAAATCCCACACCTGCTAACAGAAGCAGATACCAGTCCAGGAAGGTGGAGA 5913

Qy        814 ATATGAAAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAG 873
Db       5914 ACATGAAAGCCATAACAGCTAGACGTATTCTGCAAGGCCTAGGGCATTATTTGAAGGCAG 5973

Qy        874 AAGTGAATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGG 933
Db       5974 AAGGAAA---AGTGGAGTGCTACCGAACCCTGCATCCTGTTCCTTTGTATTCATCTAGTG 6030

Qy        934 TAATTAACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAG 993
Db       6031 TGAACCGTGCCTTTTCAAGCCCCAAGGTCGCAGTGGAAGCCTGTAACGCCATGTTGAAAG 6090

Qy        994 AGAATTACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACA 1053
Db       6091 AGAACTTTCCGACTGTGGCTTCTTACTGTATTATTCCAGAGTACGATGCCTATTTGGACA 6150
```

```
Qy       1054 TGGTGGACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAA 1113
             |||| ||||| || || |||||  |||| || |||| ||||||  || || || ||| | |
Db       6151 TGGTTGACGGAGCTTCATGCTGCTTAGACACTGCCAGTTTTTGCCCTGCAAAGCTGCGCA 6210

Qy       1114 GCTACCCAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTA 1173
             |||  |||||||| ||   ||||||| | ||     ||| |||| || || || ||  | |
Db       6211 GCTTTCCAAAGAAACACTCCTATTTGGAACCCACAATACGATCGGCAGTGCCTTCAGCGA 6270

Qy       1174 TACAGAATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCC 1233
             | ||||| ||  | || || ||  |||| || || || ||||| |||||||| || || |
Db       6271 TCCAGAACACGCTCCAGAACGTCCTGGCAGCTGCCACAAAAAGAAATTGCAATGTCACGC 6330

Qy       1234 AAATGCGAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAAT 1293
             ||||| ||||||| || || || |||||||||||| |||||||| || || ||||||||||
Db       6331 AAATGAGAGAATTGCCCGTATTGGATTCGGCGGCCTTTAATGTGGAATGCTTCAAGAAAT 6390

Qy       1294 ACGCATGCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAG 1353
             | || || ||| |||| || ||||| || |||   || ||||| ||  |||| |||  ||
Db       6391 ATGCGTGTAATAATGAATATTGGGAAACGTTTAAAGAAAACCCCATCAGGCTTACTGAAG 6450

Qy       1354 AGAACGTTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGA 1413
             | |||||     | ||  | || ||  | ||||| || ||||| || ||  | |||||||
Db       6451 AAAACGTGGTAAATTACATTACCAAATTAAAAGGACCAAAAGCTGCTGCTCTTTTTGCGA 6510

Qy       1414 ATACTCATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAA 1473
             | || |||||| | ||   |||||||||| ||||||||||    || || |||||  |||
Db       6511 AGACACATAATTTGAATATGTTGCAGGACATACCAATGGACAGGTTTGTAATGGACTTAA 6570

Qy       1474 AGAGAGATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGG 1533
             ||||||| || ||||| ||||| || || |||||||| || || ||||| ||||| ||||
Db       6571 AGAGAGACGTGAAAGTGACTCCAGGAACAAAACATACTGAAGAACGGCCCAAGGTACAGG 6630

Qy       1534 TTATTCAGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAG 1593
             | || |||||||| ||||| || || || || || || ||||| ||||| || ||  | |
Db       6631 TGATCCAGGCTGCCGATCCGCTAGCAACAGCGTATCTGTGCGGAATCCACCGAGAGCTGG 6690

Qy       1594 TCCGTAGACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGG 1653
             |  | ||| | |||||||| || || || || || |||| || || ||  ||||||| |
Db       6691 TTAGGAGATTAAATGCGGTCCTGCTTCCGAACATTCATACACTGTTTGATATGTCGGCTG 6750

Qy       1654 AAGATTTTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGG 1713
             |||| ||||| || ||||| || || || ||||| |  || ||    ||  ||||||| |
Db       6751 AAGACTTTGACGCTATTATAGCCGAGCACTTCCAGCCTGGGGATTGTGTTCTGGAAACTG 6810

Qy       1714 ACATCGCGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCC 1773
             |||||||||||||||||||||| || |||||||| || ||  |  | ||||| ||||| |
Db       6811 ACATCGCGTCGTTTGATAAAAGTGAGGACGACGCCATGGCTCTGACCGCGTTAATGATTC 6870

Qy       1774 TTGAGGACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATA 1833
             | || |||||||||||| |||  |  ||| ||     |||| |||||||| |||||| | |
Db       6871 TGGAAGACTTAGGTGTGGACGCAGAGCTGTTGACGCTGATTGAGGCGGCTTTCGGCGAAA 6930

Qy       1834 TCACATCTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCG 1893
             |  ||||  | ||  | || ||   ||| |  |||||||| || ||||||||||||| |
Db       6931 TTTCATCAATACATTTGCCCACTAAAACTAAATTTAAATTCGGAGCCATGATGAAATCTG 6990

Qy       1894 GTATGTTCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTAC 1953
             | |||||| | || |||||||| ||||||  |  | || ||  | || || |||||||
Db       6991 GAATGTTCCTCACACTGTTTGTGAACACAGTCATTAACATTGTAATCGCAAGCAGAGTGT 7050

Qy       1954 TACGTGAACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGC 2013
             |  | |||||| ||||  | || || || ||  || || || || || || || ||| |||
Db       7051 TGAGAGAACGGCTAACCGGATCACCATGTGCAGCATTCATTGGAGATGACAATATCGTGA 7110

Qy       2014 ATGGTGTCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAG 2073
             | || |||   || ||||  || ||||| || || |||||||| ||| |||| |||||||
Db       7111 AAGGAGTCAAATCGGACAAATTAATGGCAGACAGGTGCGCCACCTGGTTGAATATGGAAG 7170

Qy       2074 TAAAAATTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCC 2133
             | || ||||| ||||| ||  | ||    ||||| || || |||||||| || |||||
Db       7171 TCAAGATTATAGATGCTGTGGTGGGCGAGAAAGCGCCTTATTTCTGTGGAGGGTTTATTT 7230

Qy       2134 TGGTGGACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTA 2193
```

```
             ||  |||    | || |||||||| ||| | || |||||||| |||||||||| ||||
Db       7231 TGTGTGACTCCGTGACCGGCACAGCGTGCCGTGTGGCAGACCCCCTAAAAAGGCTGTTTA 7290

Qy       2194 AGCTTGGAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGC 2253
             ||||||| ||||| || ||| || |||   || ||     ||| |  | ||||| |||
Db       7291 AGCTTGGCAAACCTCTGGCAGCAGACGATGAACATGATGATGACAGGAGAAGGGCATTGC 7350

Qy       2254 ATGATGAAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAG 2313
             |||| ||  |||  || |||||| || | || |||       ||| |   ||||| ||||
Db       7351 ATGAAGAGTCAACACGCTGGAACCGAGTGGGTATTCTTTCAGAGCTGTGCAAGGCAGTAG 7410

Qy       2314 AATCCAGATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCG 2373
             |||| || || || | | ||    |   | | ||  | | | |  || | ||  ||||
Db       7411 AATCAAGGTATGAAACCGTAGGAACTTCCATCATAGTTATGGCCATGACTACTCTAGCTA 7470

Qy       2374 AAAGCGTTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTA 2433
               || |||||    ||||    | | |||||| ||| || || |||||||||| ||||
Db       7471 GCAGTGTTAAATCATTCAGCTACCTGAGAGGGGCCCCTATAACTCTCTACGGCTAACCTG 7530

Qy       2434 AATAGGTGACGTAGTAGAC 2452
             ||| |   |||  ||| |
Db       7531 AATGGACTACGACATAGTC 7549


RESULT 8
US-09-991-258-17
; Sequence 17, Application US/09991258
; Patent No. US20020141975A1
; GENERAL INFORMATION:
;  APPLICANT: Olmsted, Robert
;  APPLICANT:  Keith, Paula
;  APPLICANT:  Dryga, Sergey
;  APPLICANT:  Caley, Ian
;  APPLICANT:  Maughan, Maureen
;  APPLICANT:  Johnston, Robert
;  APPLICANT:  Davis, Nancy
;  APPLICANT:  Swanstrom, Ronald
;  TITLE OF INVENTION: ALPHAVIRUS VECTORS AND VIROSOMES WITH MODIFIED HIV GENES FOR USE
;  TITLE OF INVENTION:  VACCINES
;  FILE REFERENCE: 01113.0001U3
;  CURRENT APPLICATION NUMBER: US/09/991,258
;  CURRENT FILING DATE:  2001-11-16
;  PRIOR APPLICATION NUMBER: 09/902,537
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: 60/216,995
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 13584
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence; No. US20020141975A1e =
;   OTHER INFORMATION: synthetic construct
US-09-991-258-17

  Query Match           14.3%;  Score 924.6;  DB 10;  Length 13584;
  Best Local Similarity   68.0%;  Pred. No. 7e-265;
  Matches 1319;  Conservative    0;  Mismatches  614;  Indels    6;  Gaps    2;

Qy        514 AACCACCAGGAGTCGCTAGAGCGATCTCAGAAGCGGAGTTGGACGAGTACATCCGTCAAC 573
              | || ||||| ||   ||| | |||  ||  || |||||| || | || | |    ||||
Db       5617 ACCCGCCAGGCGTAAATAGGGTGATTACAAGAGAGGAGTTTGAGGCGTTCGTAGCACAAC 5676

Qy        574 ACTCCAACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTC 633
              |      | ||||||| ||| ||||| || || || || || || || || || ||||| |
Db       5677 A---ACAATGACGGTTTGATGCGGGTGCATACATCTTTTCCTCCGACACCGGTCAAGGGC 5733

Qy        634 ACCTTCAACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCG 693
              |  | ||||| |||||||||| | |||       |||   |||    | |||||  || |||
Db       5734 ATTTACAACAAAAATCAGTAAGGCAAACGGTGCTATCCGAAGTGGTGTTGGAGAGGACCG 5793
```

```
Qy       694 TCCATGAGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGA 753
Db      5794 AATTGGAGATTTCGTATGCCCCGCGCCTCGACCAAGAAAAAGAAGAATTACTACGCAAGA 5853

Qy       754 AACTGCAATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAA 813
Db      5854 AATTACAGTTAAATCCCACACCTGCTAACAGAAGCAGATACCAGTCCAGGAAGGTGGAGA 5913

Qy       814 ATATGAAAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAG 873
Db      5914 ACATGAAAGCCATAACAGCTAGACGTATTCTGCAAGGCCTAGGGCATTATTTGAAGGCAG 5973

Qy       874 AAGTGAATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGG 933
Db      5974 AAGGAAA---AGTGGAGTGCTACCGAACCCTGCATCCTGTTCCTTTGTATTCATCTAGTG 6030

Qy       934 TAATTAACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAG 993
Db      6031 TGAACCGTGCCTTTTCAAGCCCCAAGGTCGCAGTGGAAGCCTGTAACGCCATGTTGAAAG 6090

Qy       994 AGAATTACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACA 1053
Db      6091 AGAACTTTCCGACTGTGGCTTCTTACTGTATTATTCCAGAGTACGATGCCTATTTGGACA 6150

Qy      1054 TGGTGGACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAA 1113
Db      6151 TGGTTGACGGAGCTTCATGCTGCTTAGACACTGCCAGTTTTTGCCCTGCAAAGCTGCGCA 6210

Qy      1114 GCTACCCAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTA 1173
Db      6211 GCTTTCCAAAGAAACACTCCTATTTGGAACCCACAATACGATCGGCAGTGCCTTCAGCGA 6270

Qy      1174 TACAGAATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCC 1233
Db      6271 TCCAGAACACGCTCCAGAACGTCCTGGCAGCTGCCACAAAAAGAAATTGCAATGTCACGC 6330

Qy      1234 AAATGCGAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAAT 1293
Db      6331 AAATGAGAGAATTGCCCGTATTGGATTCGGCGGCCTTTAATGTGGAATGCTTCAAGAAAT 6390

Qy      1294 ACGCATGCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAG 1353
Db      6391 ATGCGTGTAATAATGAATATTGGGAAACGTTTAAAGAAAACCCCATCAGGCTTACTGAAG 6450

Qy      1354 AGAACGTTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGA 1413
Db      6451 AAAACGTGGTAAATTACATTACCAAATTAAAAGGACCAAAAGCTGCTGCTCTTTTTGCGA 6510

Qy      1414 ATACTCATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAA 1473
Db      6511 AGACACATAATTTGAATATGTTGCAGGACATACCAATGGACAGGTTTGTAATGGACTTAA 6570

Qy      1474 AGAGAGATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGG 1533
Db      6571 AGAGAGACGTGAAAGTGACTCCAGGAACAAAACATACTGAAGAACGGCCCAAGGTACAGG 6630

Qy      1534 TTATTCAGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAG 1593
Db      6631 TGATCCAGGCTGCCGATCCGCTAGCAACAGCGTATCTGTGCGGAATCCACCGAGAGCTGG 6690

Qy      1594 TCCGTAGACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGG 1653
Db      6691 TTAGGAGATTAAATGCGGTCCTGCTTCCGAACATTCATACACTGTTTGATATGTCGGCTG 6750

Qy      1654 AAGATTTTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGG 1713
Db      6751 AAGACTTTGACGCTATTATAGCCGAGCACTTCCAGCCTGGGGATTGTGTTCTGGAAACTG 6810

Qy      1714 ACATCGCGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCC 1773
Db      6811 ACATCGCGTCGTTTGATAAAAGTGAGGACGACGCCATGGCTCTGACCGCGTTAATGATTC 6870

Qy      1774 TTGAGGACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATA 1833
Db      6871 TGGAAGACTTAGGTGTGGACGCAGAGCTGTTGACGCTGATTGAGGCGGCTTTCGGCGAAA 6930

Qy      1834 TCACATCTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCG 1893
Db      6931 TTTCATCAATACATTTGCCCACTAAAACTAAATTTAAATTCGGAGCCATGATGAAATCTG 6990

Qy      1894 GTATGTTCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTAC 1953
Db      6991 GAATGTTCCTCACACTGTTTGTGAACACAGTCATTAACATTGTAATCGCAAGCAGAGTGT 7050

Qy      1954 TACGTGAACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGC 2013
Db      7051 TGAGAGAACGGCTAACCGGATCACCATGTGCAGCATTCATTGGAGATGACAATATCGTGA 7110

Qy      2014 ATGGTGTCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAG 2073
Db      7111 AAGGAGTCAAATCGGACAAATTAATGGCAGACAGGTGCGCCACCTGGTTGAATATGGAAG 7170

Qy      2074 TAAAAATTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCC 2133
Db      7171 TCAAGATTATAGATGCTGTGGTGGGCGAGAAAGCGCCTTATTTCTGTGGAGGGTTTATTT 7230

Qy      2134 TGGTGGACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTA 2193
Db      7231 TGTGTGACTCCGTGACCGGCACAGCGTGCCGTGTGGCAGACCCCCTAAAAAGGCTGTTTA 7290

Qy      2194 AGCTTGGAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGC 2253
Db      7291 AGCTTGGCAAACCTCTGGCAGCAGACGATGAACATGATGATGACAGGAGAAGGGCATTGC 7350

Qy      2254 ATGATGAAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAG 2313
Db      7351 ATGAAGAGTCAACACGCTGGAACCGAGTGGGTATTCTTTCAGAGCTGTGCAAGGCAGTAG 7410

Qy      2314 AATCCAGATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCG 2373
Db      7411 AATCAAGGTATGAAACCGTAGGAACTTCCATCATAGTTATGGCCATGACTACTCTAGCTA 7470

Qy      2374 AAAGCGTTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTA 2433
Db      7471 GCAGTGTTAAATCATTCAGCTACCTGAGAGGGGCCCCTATAACTCTCTACGGCTAACCTG 7530

Qy      2434 AATAGGTGACGTAGTAGAC 2452
Db      7531 AATGGACTACGACATAGTC 7549
```

RESULT 9
US-09-991-258-2
; Sequence 2, Application US/09991258
; Patent No. US20020141975A1
; GENERAL INFORMATION:
; APPLICANT: Olmsted, Robert
; APPLICANT: Keith, Paula
; APPLICANT: Dryga, Sergey
; APPLICANT: Caley, Ian
; APPLICANT: Maughan, Maureen
; APPLICANT: Johnston, Robert
; APPLICANT: Davis, Nancy
; APPLICANT: Swanstrom, Ronald
; TITLE OF INVENTION: ALPHAVIRUS VECTORS AND VIROSOMES WITH MODIFIED HIV GENES FOR USE I
; TITLE OF INVENTION: VACCINES
; FILE REFERENCE: 01113.0001U3
; CURRENT APPLICATION NUMBER: US/09/991,258
; CURRENT FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: 09/902,537
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: 60/216,995
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2

```
;   LENGTH: 7479
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence; No. US20020141975A1e =
;   OTHER INFORMATION: synthetic construct
;   NAME/KEY: CDS
;   LOCATION: (1)...(7479)
US-09-991-258-2

  Query Match           14.1%;  Score 916.2;  DB 10;  Length 7479;
  Best Local Similarity  68.2%;  Pred. No. 1.5e-262;
  Matches 1304;  Conservative    0;  Mismatches  603;  Indels    6;  Gaps    2;

Qy       514 AACCACCAGGAGTCGCTAGAGCGATCTCAGAAGCGGAGTTGGACGAGTACATCCGTCAAC 573
Db      5573 ACCCGCCAGGCGTAAATAGGGTGATTACAAGAGAGGAGTTTGAGGCGTTCGTAGCACAAC 5632

Qy       574 ACTCCAACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTC 633
Db      5633 A---ACAATGACGGTTTGATGCGGGTGCATACATCTTTTCCTCCGACACCGGTCAAGGGC 5689

Qy       634 ACCTTCAACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCG 693
Db      5690 ATTTACAACAAAAATCAGTAAGGCAAACGGTGCTATCCGAAGTGGTGTTGGAGAGGACCG 5749

Qy       694 TCCATGAGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGA 753
Db      5750 AATTGGAGATTTCGTATGCCCCGCGCCTCGACCAAGAAAAAGAAGAATTACTACGCAAGA 5809

Qy       754 AACTGCAATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAA 813
Db      5810 AATTACAGTTAAATCCCACACCTGCTAACAGAAGCAGATACCAGTCCAGGAAGGTGGAGA 5869

Qy       814 ATATGAAAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAG 873
Db      5870 ACATGAAAGCCATAACAGCTAGACGTATTCTGCAAGGCCTAGGGCATTATTTGAAGGCAG 5929

Qy       874 AAGTGAATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGG 933
Db      5930 AAGGAAA---AGTGGAGTGCTACCGAACCCTGCATCCTGTTCCTTTGTATTCATCTAGTG 5986

Qy       934 TAATTAACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAG 993
Db      5987 TGAACCGTGCCTTTTCAAGCCCCAAGGTCGCAGTGGAAGCCTGTAACGCCATGTTGAAAG 6046

Qy       994 AGAATTACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACA 1053
Db      6047 AGAACTTTCCGACTGTGGCTTCTTACTGTATTATTCCAGAGTACGATGCCTATTTGGACA 6106

Qy      1054 TGGTGGACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAA 1113
Db      6107 TGGTTGACGGAGCTTCATGCTGCTTAGACACTGCCAGTTTTTGCCCTGCAAAGCTGCGCA 6166

Qy      1114 GCTACCCAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTA 1173
Db      6167 GCTTTCCAAAGAAACACTCCTATTTGGAACCCACAATACGATCGGCAGTGCCTTCAGCGA 6226

Qy      1174 TACAGAATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCC 1233
Db      6227 TCCAGAACACGCTCCAGAACGTCCTGGCAGCTGCCACAAAAAGAAATTGCAATGTCACGC 6286

Qy      1234 AAATGCGAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAAT 1293
Db      6287 AAATGAGAGAATTGCCCGTATTGGATTCGGCGGCCTTTAATGTGGAATGCTTCAAGAAAT 6346

Qy      1294 ACGCATGCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAG 1353
Db      6347 ATGCGTGTAATAATGAATATTGGGAAACGTTTAAAGAAAACCCCATCAGGCTTACTGAAG 6406

Qy      1354 AGAACGTTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGA 1413
Db      6407 AAAACGTGGTAAATTACATTACCAAATTAAAAGGACCAAAAGCTGCTGCTCTTTTTGCGA 6466

Qy      1414 ATACTCATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAA 1473
Db      6467 AGACACATAATTTGAATATGTTGCAGGACATACCAATGGACAGGTTTGTAATGGACTTAA 6526

Qy      1474 AGAGAGATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGG 1533
Db      6527 AGAGAGACGTGAAAGTGACTCCAGGAACAAAACATACTGAAGAACGGCCCAAGGTACAGG 6586

Qy      1534 TTATTCAGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAG 1593
Db      6587 TGATCCAGGCTGCCGATCCGCTAGCAACAGCGTATCTGTGCGGAATCCACCGAGAGCTGG 6646

Qy      1594 TCCGTAGACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGG 1653
Db      6647 TTAGGAGATTAAATGCGGTCCTGCTTCCGAACATTCATACACTGTTTGATATGTCGGCTG 6706

Qy      1654 AAGATTTTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGG 1713
Db      6707 AAGACTTTGACGCTATTATAGCCGAGCACTTCCAGCCTGGGGATTGTGTTCTGGAAACTG 6766

Qy      1714 ACATCGCGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCC 1773
Db      6767 ACATCGCGTCGTTTGATAAAAGTGAGGACGACGCCATGGCTCTGACCGCGTTAATGATTC 6826

Qy      1774 TTGAGGACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATA 1833
Db      6827 TGGAAGACTTAGGTGTGGACGCAGAGCTGTTGACGCTGATTGAGGCGGCTTTCGGCGAAA 6886

Qy      1834 TCACATCTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCG 1893
Db      6887 TTTCATCAATACATTTGCCCACTAAAACTAAATTTAAATTCGGAGCCATGATGAAATCTG 6946

Qy      1894 GTATGTTCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTAC 1953
Db      6947 GAATGTTCCTCACACTGTTTGTGAACACAGTCATTAACATTGTAATCGCAAGCAGAGTGT 7006

Qy      1954 TACGTGAACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGC 2013
Db      7007 TGAGAGAACGGCTAACCGGATCACCATGTGCAGCATTCATTGGAGATGACAATATCGTGA 7066

Qy      2014 ATGGTGTCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAG 2073
Db      7067 AAGGAGTCAAATCGGACAAATTAATGGCAGACAGGTGCGCCACCTGGTTGAATATGGAAG 7126

Qy      2074 TAAAAATTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCC 2133
Db      7127 TCAAGATTATAGATGCTGTGGTGGGCGAGAAAGCGCCCTATTTCTGTGGAGGGTTTATTT 7186

Qy      2134 TGGTGGACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTA 2193
Db      7187 TGTGTGACTCCGTGACCGGCACAGCGTGCCGTGTGGCAGACCCCCTAAAAAGGCTGTTTA 7246

Qy      2194 AGCTTGGAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGC 2253
Db      7247 AGCTTGGCAAACCTCTGGCAGCAGACGATGAACATGATGATGACAGGAGAAGGGCATTGC 7306

Qy      2254 ATGATGAAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAG 2313
Db      7307 ATGAAGAGTCAACACGCTGGAACCGAGTGGGTATTCTTTCAGAGCTGTGCAAGGCAGTAG 7366

Qy      2314 AATCCAGATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCG 2373
Db      7367 AATCAAGGTATGAAACCGTAGGAACTTCCATCATAGTTATGGCCATGACTACTCTAGCTA 7426

Qy      2374 AAAGCGTTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGC 2426
Db      7427 GCAGTGTTAAATCATTCAGCTACCTGAGAGGGGCCCCTATAACTCTCTACGGC 7479


RESULT 10
US-09-507-362-102
; Sequence 102, Application US/09507362
```

```
; Publication No. US20030096397A1
;   GENERAL INFORMATION:
;        APPLICANT: Dubensky Jr., Thomas W.
;                   Polo, John M.
;                   Belli, Barbara A.
;                   Schlesinger, Sondra
;                   Dryga, Sergey A.
;                   Frolov, Ilya
;        TITLE OF INVENTION: RECOMBINANT ALPHAVIRUS-BASED VECTORS
;                            WITH REDUCED INHIBITION OF CELLULAR MACRO-MOLECULAR
;                            SYNTHESIS
;        NUMBER OF SEQUENCES: 125
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Seed Intellectual Property Law Group PLLC
;             STREET: 701 Fifth Avenue, Suite 6300
;             CITY: Seattle
;             STATE: Washington
;             COUNTRY: USA
;             ZIP: 98104-7092
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/507,362
;             FILING DATE: 18-Feb-2000
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McMasters, David D.
;             REGISTRATION NUMBER: 33,963
;             REFERENCE/DOCKET NUMBER: 930049.457D6 /1196.011
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (206) 622-4900
;             TELEFAX: (206) 682-6031
;   INFORMATION FOR SEQ ID NO: 102:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 8000 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-09-507-362-102

  Query Match             12.8%;  Score 827.6;  DB 11;  Length 8000;
  Best Local Similarity   61.9%;  Pred. No. 5.9e-236;
  Matches 1390;  Conservative    0;  Mismatches  829;  Indels   25;  Gaps    4;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
              ||||||     |  |  || | ||| || || ||  |||| |||||||  || ||| | |
Db       5746 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5805

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
              || |||| || || |  ||     | ||   |||||| |  ||||| ||    ||||  |
Db       5806 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTCCTGG 5865

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
              | |    | | ||||||   |||||       | |||| ||    |  |    |    |
Db       5866 AAAGAATTCATGCCCCGGTGCTCGACACGTCGAAAGAGGAACAACTCAAACTCAGGTACC 5925

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
              |  |      || | ||||  || | ||| ||||| || || || ||||||||||||  ||
Db       5926 AGATGATGCCCACCGAAGCCAACAAAAGTAGGTACCAGTCTCGTAAAGTAGAAAATCAGA 5985

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
              |||| || ||  | |||||||| | || ||| |  |    |||  ||  |   ||
Db       5986 AAGCCATAACCACTGAGCGACTACTGTCAGGACTACGACTGTATAACTCTGCCACAGATC 6045

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
              | ||    || || || |   |||  |||||   |||| | |||||    |  |||
Db       6046 AGCCA---GAATGCTATAAGATCACCTATCCGAAACCATTGTACTCCAGTAGCGTACCGG 6102

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
```

```
                |  |   |  | || || |||| ||       || |||     | | || |||||  |
Db       6103 CGAACTACTCCGATCCACAGTTCGCTGTAGCTGTCTGTAACAACTATCTGCATGAGAACT 6162

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
              | || ||||||||    ||||   || || || || |||||||| ||  | ||| |||| |
Db       6163 ATCCGACAGTAGCATCTTATCAGATTACTGACGAGTACGATGCTTACTTGGATATGGTAG 6222

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
              ||||  ||     ||| || ||||| || || || || || ||||| || ||||| ||||
Db       6223 ACGGGACAGTCGCCTGCCTGGATACTGCAACCTTCTGCCCCGCTAAGCTTAGAAGTTACC 6282

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
              | || || |||   |||      ||  | ||  |    || || |||||  | || ||||
Db       6283 CGAAAAAACATGAGTATAGAGCCCCGAATATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6342

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
              | ||  ||||||||||    |   ||| || |||||||| ||||||||||| || ||||
Db       6343 ACACGCTACAAAATGTGCTCATTGCCGCAACTAAAAGAAATTGCAACGTCACGCAGATGC 6402

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
              | ||| | ||     | || || ||| |||| ||||| || || ||     ||||| ||||
Db       6403 GTGAACTGCCAACACTGGACTCAGCGACATTCAATGTCGAATGCTTTCGAAAATATGCAT 6462

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
              | ||||| ||||| |||||     ||        || || ||| || | || || |||   |
Db       6463 GTAATGACGAGTATTGGGAGGAGTTCGCTCGGAAGCCAATTAGGATTACCACTGAGTTTG 6522

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
              | ||   ||||||  | |  |||||||| || || || || ||| | ||||| || ||
Db       6523 TCACCGCATATGTAGCTAGACTGAAAGGCCCTAAGGCCGCCGCACTATTTGCAAAGACGT 6582

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
              |||||  |    || ||||| ||  | || ||||||  |||||||||||  | || ||||
Db       6583 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6642

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
              | || |||||||| || |||||||||| ||||| ||  | || || || || || || | |
Db       6643 ACGTGAAAGTTACACCAGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6702

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
              | ||  |||||  ||||| || ||||||  | |||||||| || ||||||||||| ||||
Db       6703 AAGCCGCAGAACCCCTGGCGACTGCTTACTTATGCGGGATTCACCGGGAATTAGTGCGTA 6762

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
              | || |  || ||  | || ||||| || || || || || |||||||| |||| ||||
Db       6763 GGCTTACGGCCGTCTTGCTTCCAAACATTCACACGCTTTTTGACATGTCGGCGGAGGATT 6822

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
              ||||||| || || || ||||| ||| | || |||||||| ||| |||| ||||| ||||
Db       6823 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGCGACCCGGTACTGGAGACGGATATCG 6882

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
              | || || || |||||| |||||||||||| ||  |  | |  |||||||| | ||||
Db       6883 CATCATTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGTCTGATGATCTTGGAGG 6942

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
              || | ||||| || ||||| ||  | || ||||| |||   || || ||  | ||  |||
Db       6943 ACCTGGGTGTGGATCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 7002

Qy       1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
              |    || |||||||| || ||  | |||||||| || || ||||||||||||| ||||
Db       7003 CCACCCATCTACCTACGGGTACTCGTTTTAAATTCGGGGCGATGATGAAATCCGGAATGT 7062

Qy       1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
              || | || || |||||||||||| |  | ||| || | || || ||||||||||||  |
Db       7063 TCCTCACACTTTTTGTCAACACAGTTTTGAATGTCGTTATCGCCAGCAGAGTACTAGAAG 7122

Qy       1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
              | ||| | |  |||||     || || || |  || |||||||| |||||  | ||||| |
Db       7123 AGCGGCTTAAAACGTCCAGATGTGCAGCGTTCATTGGCGACGACAACATCATACATGGAG 7182

Qy       2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
              | || || ||||     ||||| ||||| |||||||| ||||| |||||||| || || |
```

```
Db       7183 TAGTATCTGACAAAGAAATGGCTGAGAGGTGCGCCACCTGGCTCAACATGGAGGTTAAGA 7242

Qy       2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139

Db       7243 TCATCGACGCAGTCATCGGTGAGAGACCACCTTACTTCTGCGGCGGATTTATCTTGCAAG 7302

Qy       2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199

Db       7303 ATTCGGTTACTTCCACAGCGTGCCGCGTGGCGGATCCCCTGAAAAGGCTGTTTAAGTTGG 7362

Qy       2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259

Db       7363 GTAAACCGCTCCCAGCCGACGACGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 7422

Qy       2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319

Db       7423 AAACAAAGGCGTGGTTTAGAGTAGGTATAACAGGCACTTTAGCAGTGGCCGTGACGACCC 7482

Qy       2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379

Db       7483 GGTATGAGGTAGACAATATTACACCTGTCCTACTGGCATTGAGAACTTTTGCCCAGAGCA 7542

Qy       2380 TTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGG 2439

Db       7543 AAAGAGCATTCCAAGCCATCAGAGGGGAAATAAAGCATCTCTACGGTGGTCCTAAATAGT 7602

Qy       2440 TGACGTAGTAGACACGCACCTAC----------CCACCGGCAGAATGTTTCCATACC 2486

Db       7603 CAGCATAGTACATTTCATCTGACTAATACTACAACACCACCACCATGAATAGAGGATTCT 7662

Qy       2487 CTCAGCTGAACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACC 2546

Db       7663 TTAACATGCTCGGCCGCCGCCCCTTCCCGGCCCCCACTGCCATGTGGAGGCCGCGGAGAA 7722

Qy       2547 CTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTA 2606

Db       7723 GGAGGCAGGCGGCCCCGATGCCTGCCCGCAACGGGCTGGCTTCTCAAATCCAGCAACTGA 7782

Qy       2607 GGAGGTCGATAGTCAACTTGACTTTCAAACAACGATCACCTAATCCGCCGCCAGGTCCA- 2665

Db       7783 CCACAGCCGTCAGTGCCCTAGTCATTGGACAGGCAACTAGACCTCAACCCCCACGTCCAC 7842

Qy       2666 -----CCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGA 2720

Db       7843 GCCCGCCACCGCGCCAGAAGAAGCAGGCGCCCAAGCAACCACCGAAGCCGAAGAAACCAA 7902

Qy       2721 AGAAGCAGCAAGCCAAGAGGACGAAACGC---AAGCCTAAACCAGGGAAACGACAACGTA 2777

Db       7903 AAACGCAGGAGAAGAAGAAGAAGCAACCTGCAAAACCCAAACCCGGAAAGAGACAGCGCA 7962

Qy       2778 TGTGTATGAAGTTGGAGTCGGACA 2801

Db       7963 TGGCACTTAAGTTGGAGGCCGACA 7986


RESULT 11
US-09-507-362-101
; Sequence 101, Application US/09507362
; Publication No. US20030096397A1
;  GENERAL INFORMATION:
;       APPLICANT: Dubensky Jr., Thomas W.
;                  Polo, John M.
;                  Belli, Barbara A.
;                  Schlesinger, Sondra
;                  Dryga, Sergey A.
;                  Frolov, Ilya
;       TITLE OF INVENTION: RECOMBINANT ALPHAVIRUS-BASED VECTORS
;                           WITH REDUCED INHIBITION OF CELLULAR MACRO-MOLECULAR
;                           SYNTHESIS
;       NUMBER OF SEQUENCES: 125
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Seed Intellectual Property Law Group PLLC
;            STREET: 701 Fifth Avenue, Suite 6300
;            CITY: Seattle
;            STATE: Washington
;            COUNTRY: USA
;            ZIP: 98104-7092
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/507,362
;            FILING DATE: 18-Feb-2000
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: McMasters, David D.
;            REGISTRATION NUMBER: 33,963
;            REFERENCE/DOCKET NUMBER: 930049.457D6 /1196.011
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (206) 622-4900
;            TELEFAX: (206) 682-6031
;  INFORMATION FOR SEQ ID NO: 101:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 8000 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 101:
US-09-507-362-101

  Query Match             12.7%;  Score 824.4;  DB 11;  Length 8000;
  Best Local Similarity   61.9%;  Pred. No. 5.4e-235;
  Matches 1388;  Conservative    0;  Mismatches  831;  Indels   25;  Gaps    4;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639

Db       5746 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5805

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699

Db       5806 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCACAATGTCCTGG 5865

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759

Db       5866 AAAGAATTCATGCCCCGGTGCTCGACACGTCGAAAGAGGAACAACTCAAACTCAGGTACC 5925

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819

Db       5926 AGATGATGCCCACCGAAGCCAACAAAAGTAGGTACCAGTCTCGTAAAGTAGAAAATCAGA 5985

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879

Db       5986 AAGCCATAACCACTGAGCGACTACTGTCAGGACTACGACTGTATAACTCTGCCACAGATC 6045

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939

Db       6046 AGCCA---GAATGCTATAAGATCACCTATCCGAAACCATTGTACTCCAGTAGCGTACCGG 6102

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999

Db       6103 CGAACTACTCCGATCCACAGTTCGCTGTAGCTGTCTGTAACAACTATCTGCATGAGAACT 6162

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059

Db       6163 ATCCGACAGTAGCATCTTATCAGATTACTGACGAGTACGATGCTTACTTGGATATGGTAG 6222

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119

Db       6223 ACGGGACAGTCGCCTGCCTGGATACTGCAACCTTCTGCCCCGCTAAGCTTAGAAGTTACC 6282

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179

Db       6283 CGAAAAAACATGAGTATAGAGCCCCGAATATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6342
```

```
Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
Db       6343 ACACGCTACAAAATGTGCTCATTGCCGCAACTAAAAGAAATTGCAACGTCACGCAGATGC 6402

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
Db       6403 GTGAACTGCCAACACTGGACTCAGCGACATTCAATGTCGAATGCTTTCGAAAATATGCAT 6462

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
Db       6463 GTAATGACGAGTATTGGGAGGAGTTCGCTCGGAAGCCAATTAGGATTACCACTGAGTTTG 6522

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
Db       6523 TCACCGCATATGTAGCTAGACTGAAAGGCCCTAAGGCCGCCGCACTATTTGCAAAGACGT 6582

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
Db       6583 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6642

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
Db       6643 ACGTGAAAGTTACACCAGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6702

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
Db       6703 AAGCCGCAGAACCCCTGGCGACTGCTTACTTATGCGGGATTCACCGGGAATTAGTGCGTA 6762

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
Db       6763 GGCTTACGGCCGTCTTGCTTCCAAACATTCACACGCTTTTTGACATGTCGGCGGAGGATT 6822

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
Db       6823 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGCGACCCGGTACTGGAGACGGATATCG 6882

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
Db       6883 CATCATTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGTCTGATGATCTTGGAGG 6942

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
Db       6943 ACCTGGGTGTGGATCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 7002

Qy       1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
Db       7003 CCACCCATCTACCTACGGGTACTCGTTTTAAATTCGGGGCGATGATGAAATCCGGAATGT 7062

Qy       1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
Db       7063 TCCTCACACTTTTTGTCAACACAGTTTTGAATGTCGTTATCGCCAGCAGAGTACTAGAAG 7122

Qy       1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
Db       7123 AGCGGCTTAAAACGTCCAGATGTGCAGCGTTCATTGGCGACGACAACATCATACATGGAG 7182

Qy       2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
Db       7183 TAGTATCTGACAAAGAAATGGCTGAGAGGTGCGCCACCTGGCTCAACATGGAGGTTAAGA 7242

Qy       2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139
Db       7243 TCATCGACGCAGTCATCGGTGAGAGACCACCTTACTTCTGCGGCGGATTTATCTTGCAAG 7302

Qy       2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199
Db       7303 ATTCGGTTACTTCCACAGCGTGCCGCGTGGCGGATCCCCTGAAAAGGCTGTTTAAGTTGG 7362

Qy       2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
Db       7363 GTAAACCGCTCCCAGCCGACGACGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 7422

Qy       2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319

Db       7423 AAACAAAGGCGTGGTTTAGAGTAGGTATAACAGGCACTTTAGCAGTGGCCGTGACGACCC 7482

Qy       2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379
Db       7483 GGTATGAGGTAGACAATATTACACCTGTCCTACTGGCATTGAGAACTTTTGCCCAGAGCA 7542

Qy       2380 TTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGG 2439
Db       7543 AAAGAGCATTCCAAGCCATCAGAGGGGAAATAAAGCATCTCTACGGTGGTCCTAAATAGT 7602

Qy       2440 TGACGTAGTAGACACGCACCTAC----------CCACCGGCAGAATGTTTCCATACC 2486
Db       7603 CAGCATAGTTCATTTCATCTGACTAATACTACAACACCACCACCATGAATAGAGGATTCT 7662

Qy       2487 CTCAGCTGAACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACC 2546
Db       7663 TTAACATGCTCGGCCGCCGCCCCTTCCCGGCCCCCACTGCCATGTGGAGGCCGCGGAGAA 7722

Qy       2547 CTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTA 2606
Db       7723 GGAGGCAGGCGGCCCCGATGCCTGCCCGCAACGGGCTGGCTTCTCAAATCCAGCAACTGA 7782

Qy       2607 GGAGGTCGATAGTCAACTTGACTTTCAAACAACGATCACCTAATCCGCCGCCAGGTCCA- 2665
Db       7783 CCACAGCCGTCAGTGCCCTAGTCATTGGACAGGCAACTAGACCTCAACCCCCATGTCCAC 7842

Qy       2666 -----CCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGA 2720
Db       7843 GCCCGCCACCGCGCCAGAAGAAGCAGGCGCCCAAGCAACCACCGAAGCCGAAGAAACCAA 7902

Qy       2721 AGAAGCAGCAAGCCAAGAGGACGAAACGC---AAGCCTAAACCAGGGAAACGACAACGTA 2777
Db       7903 AAACGCAGGAGAAGAAGAAGAAGCAACCTGCAAAACCCAAACCCGGAAAGAGACAGCGCA 7962

Qy       2778 TGTGTATGAAGTTGGAGTCGGACA 2801
Db       7963 TGGCACTTAAGTTGGAGGCCGACA 7986
```

RESULT 12
US-09-733-042-1
; Sequence 1, Application US/09733042
; Patent No. US20020168709A1
; GENERAL INFORMATION:
; APPLICANT: Hennecke, Frank
; APPLICANT: Renner, Wolfgang A.
; TITLE OF INVENTION: Replicon Based Activation of Endogenous Genes
; FILE REFERENCE: 1700.0100001
; CURRENT APPLICATION NUMBER: US/09/733,042
; CURRENT FILING DATE: 2000-12-11
; PRIOR APPLICATION NUMBER: US 60/169,988
; PRIOR FILING DATE: 1999-12-10
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 11282
; TYPE: DNA
; ORGANISM: pCYTts
US-09-733-042-1

Query Match 12.6%; Score 816; DB 10; Length 11282;
Best Local Similarity 65.1%; Pred. No. 2.3e-232;
Matches 1219; Conservative 0; Mismatches 650; Indels 3; Gaps 1;

```
Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
Db       6638 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 6697

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
Db       6698 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTCCTGG 6757
```

```
Qy       700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
Db      6758 AAAGAATTCATGCCCCGGTGCTCGACACGTCGAAAGAGGAACAACTCAAACTCAGGTACC 6817

Qy       760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
Db      6818 AGATGATGCCCACCGAAGCCAACAAAAGTAGGTACCAGTCTCGTAAAGTAGAAAATCAGA 6877

Qy       820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
Db      6878 AAGCCATAACCACTGAGCGACTACTGTCAGGACTACGACTGTATAACTCTGCCACAG--- 6934

Qy       880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
Db      6935 ATCAGCCAGAATGCTATAAGATCACCTATCCGAAACCATTGTACTCCAGTAGCGTACCGG 6994

Qy       940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
Db      6995 CGAACTACTCCGATCCACAGTTCGCTGTAGCTGTCTGTAACAACTATCTGCATGAGAACT 7054

Qy      1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
Db      7055 ATCCGACAGTAGCATCTTATCAGATTACTGACGAGTACGATGCTTACTTGGATATGGTAG 7114

Qy      1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
Db      7115 ACGAGACAGTCGCATGCCTGGATACTGCAACCTTCTGCCCCGCTAAGCTTAGAAGTTACC 7174

Qy      1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
Db      7175 CGAAAAAACATGAGTATAGAGCCCCGAATATCCGCAGTGCGGTTCCATCAGCGATGCAGA 7234

Qy      1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
Db      7235 ACACGCTACAAAATGTGCTCATTGCCGCAACTAAAAGAAATTGCAACGTCACGCAGATGC 7294

Qy      1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
Db      7295 GTGAACTGCCAACACTGGACTCAGCGACATTCAATGTCGAATGCTTTCGAAAATATGCAT 7354

Qy      1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
Db      7355 GTAATGACGAGTATTGGGAGGAGTTCGCTCGGAAGCCAATTAGGATTACCACTGAGTTTG 7414

Qy      1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
Db      7415 TCACCGCATATGTAGCTAGACTGAAAGGCCCTAAGGCCGCCGCACTATTTGCAAAGACGT 7474

Qy      1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
Db      7475 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 7534

Qy      1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
Db      7535 ACGTGAAAGTTACACCAGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 7594

Qy      1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
Db      7595 AAGCCGCAGAACCCCTGGCGACTGCTTACTTATGCGGGATTCACCGGGAATTAGTGCGTA 7654

Qy      1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
Db      7655 GGCTTACGGCCGTCTTGCTTCCAAACATTCACACGCTTTTTGACATGTCGGCGGAGGATT 7714

Qy      1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
Db      7715 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGCGACCCGGTACTGGAGACGGATATCG 7774

Qy      1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
Db      7775 CATCATTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGTCTGATGATCTTGGAGG 7834

Qy      1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
```

```
Db      7835 ACCTGGGTGTGGATCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 7894

Qy      1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
Db      7895 CCACCCATCTACCTACGGGTACTCGTTTTAAATTCGGGGCGATGATGAAATCCGGAATGT 7954

Qy      1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
Db      7955 TCCTCACACTTTTTGTCAACACAGTTTTGAATGTCGTTATCGCCAGCAGAGTACTAGAAG 8014

Qy      1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
Db      8015 AGCGGCTTAAAACGTCCAGATGTGCAGCGTTCATTGGCGACGACAACATCATACATGGAG 8074

Qy      2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
Db      8075 TAGTATCTGACAAAGAAATGGCTGAGAGGTGCGCCACCTGGCTCAACATGGAGGTTAAGA 8134

Qy      2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139
Db      8135 TCATCGACGCAGTCATCGGTGAGAGACCACCTTACTTCTGCGGCGGATTTATCTTGCAAG 8194

Qy      2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199
Db      8195 ATTCGGTTACTTCCACAGCGTGCCGCGTGGCGGATCCCCTGAAAAGGCTGTTTAAGTTGG 8254

Qy      2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
Db      8255 GTAAACCGCTCCCAGCCGACGACGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 8314

Qy      2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319
Db      8315 AAACAAAGGCGTGGTTTAGAGTAGGTATAACAGGCACTTTAGCAGTGGCCGTGACGACCC 8374

Qy      2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379
Db      8375 GGTATGAGGTAGACAATATTACACCTGTCCTACTGGCATTGAGAACTTTTGCCCAGAGCA 8434

Qy      2380 TTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGG 2439
Db      8435 AAAGAGCATTCCAAGCCATCAGAGGGGAAATAAAGCATCTCTACGGTGGTCCTAAATAGT 8494

Qy      2440 TGACGTAGTAGA 2451
Db      8495 CAGCATAGTACA 8506
```

RESULT 13
US-09-275-883-1
; Sequence 1, Application US/09275883
; Publication No. US20030053988A1
; GENERAL INFORMATION:
; APPLICANT: Renner, Wolfgang A.
; APPLICANT: Nieba, Lars
; APPLICANT: Boorsma, Marco
; TITLE OF INVENTION: Inducible Alphaviral Gene Expression System
; FILE REFERENCE: 1700.0020001
; CURRENT APPLICATION NUMBER: US/09/275,883
; CURRENT FILING DATE: 1999-03-25
; EARLIER APPLICATION NUMBER: US 60/079,562
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 11282
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:cDNA
US-09-275-883-1

Query Match 12.6%; Score 816; DB 11; Length 11282;

Best Local Similarity 65.1%; Pred. No. 2.3e-232;
Matches 1219; Conservative 0; Mismatches 650; Indels 3; Gaps 1;

```
Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
Db       6638 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 6697

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
Db       6698 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTCCTGG 6757

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
Db       6758 AAAGAATTCATGCCCCGGTGCTCGACACGTCGAAAGAGGAACAACTCAAACTCAGGTACC 6817

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
Db       6818 AGATGATGCCCACCGAAGCCAACAAAAGTAGGTACCAGTCTCGTAAAGTAGAAAATCAGA 6877

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
Db       6878 AAGCCATAACCACTGAGCGACTACTGTCAGGACTACGACTGTATAACTCTGCCACAG--- 6934

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
Db       6935 ATCAGCCAGAATGCTATAAGATCACCTATCCGAAACCATTGTACTCCAGTAGCGTACCGG 6994

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
Db       6995 CGAACTACTCCGATCCACAGTTCGCTGTAGCTGTCTGTAACAACTATCTGCATGAGAACT 7054

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
Db       7055 ATCCGACAGTAGCATCTTATCAGATTACTGACGAGTACGATGCTTACTTGGATATGGTAG 7114

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
Db       7115 ACGAGACAGTCGCATGCCTGGATACTGCAACCTTCTGCCCCGCTAAGCTTAGAAGTTACC 7174

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
Db       7175 CGAAAAAACATGAGTATAGAGCCCCGAATATCCGCAGTGCGGTTCCATCAGCGATGCAGA 7234

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
Db       7235 ACACGCTACAAAATGTGCTCATTGCCGCAACTAAAAGAAATTGCAACGTCACGCAGATGC 7294

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
Db       7295 GTGAACTGCCAACACTGGACTCAGCGACATTCAATGTCGAATGCTTTCGAAAATATGCAT 7354

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
Db       7355 GTAATGACGAGTATTGGGAGGAGTTCGCTCGGAAGCCAATTAGGATTACCACTGAGTTTG 7414

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
Db       7415 TCACCGCATATGTAGCTAGACTGAAAGGCCCTAAGGCCGCCGCACTATTTGCAAAGACGT 7474

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
Db       7475 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 7534

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
Db       7535 ACGTGAAAGTTACACCAGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 7594

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
Db       7595 AAGCCGCAGAACCCCTGGCGACTGCTTACTTATGCGGGATTCACCGGGAATTAGTGCGTA 7654

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
Db       7655 GGCTTACGGCCGTCTTGCTTCCAAACATTCACACGCTTTTTGACATGTCGGCGGAGGATT 7714

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
Db       7715 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGCGACCCGGTACTGGAGACGGATATCG 7774

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
Db       7775 CATCATTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGTCTGATGATCTTGGAGG 7834

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
Db       7835 ACCTGGGTGTGGATCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 7894

Qy       1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
Db       7895 CCACCCATCTACCTACGGGTACTCGTTTTAAATTCGGGGCGATGATGAAATCCGGAATGT 7954

Qy       1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
Db       7955 TCCTCACACTTTTTGTCAACACAGTTTTGAATGTCGTTATCGCCAGCAGAGTACTAGAAG 8014

Qy       1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
Db       8015 AGCGGCTTAAAACGTCCAGATGTGCAGCGTTCATTGGCGACGACAACATCATACATGGAG 8074

Qy       2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
Db       8075 TAGTATCTGACAAAGAAATGGCTGAGAGGTGCGCCACCTGGCTCAACATGGAGGTTAAGA 8134

Qy       2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139
Db       8135 TCATCGACGCAGTCATCGGTGAGAGACCACCTTACTTCTGCGGCGGATTTATCTTGCAAG 8194

Qy       2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199
Db       8195 ATTCGGTTACTTCCACAGCGTGCCGCGTGGCGGATCCCCTGAAAAGGCTGTTTAAGTTGG 8254

Qy       2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
Db       8255 GTAAACCGCTCCCAGCCGACGACGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 8314

Qy       2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319
Db       8315 AAACAAAGGCGTGGTTTAGAGTAGGTATAACAGGCACTTTAGCAGTGGCCGTGACGACCC 8374

Qy       2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379
Db       8375 GGTATGAGGTAGACAATATTACACCTGTCCTACTGGCATTGAGAACTTTTGCCCAGAGCA 8434

Qy       2380 TTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGG 2439
Db       8435 AAAGAGCATTCCAAGCCATCAGAGGGGAAATAAAGCATCTCTACGGTGGTCCTAAATAGT 8494

Qy       2440 TGACGTAGTAGA 2451
Db       8495 CAGCATAGTACA 8506
```

RESULT 14
US-09-190-246-4
; Sequence 4, Application US/09190246
; Publication No. US20030180257A1
; GENERAL INFORMATION:
; APPLICANT: Parrington, Mark
; APPLICANT: Li, Xiaomao
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: ALPHAVIRUS VECTORS FOR PARAMYXOVIRUS VACCINES
; FILE REFERENCE: Parrington et al.
; CURRENT APPLICATION NUMBER: US/09/190,246
; CURRENT FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 4
; LENGTH: 8100
; TYPE: DNA
; ORGANISM: Semliki Forest virus
US-09-190-246-4

Query Match 11.8%; Score 764; DB 12; Length 8100;
Best Local Similarity 63.2%; Pred. No. 7.2e-217;
Matches 1175; Conservative 0; Mismatches 685; Indels 0; Gaps 0;

```
Qy       594 GCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTCAACAGAAATCAGTA 653
Db      5251 GCGGGTGCATATATTTTCTCCTCGGACACTGGCAGCGGACATTTACAACAAAAATCCGTT 5310

Qy       654 CGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATGAGAAGTATTACGCC 713
Db      5311 AGGCAGCACAATCTCCAGTGCGCACAACTGGATGCGGTCCAGGAGGAGAAAATGTACCCG 5370

Qy       714 CCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGCAATTATGCGCCTCT 773
Db      5371 CCAAAATTGGATACTGAGAGGGAGAAGCTGTTGCTGCTGAAAATGCAGATGCACCCATCG 5430

Qy       774 GAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGAAAGCAATTACAGCG 833
Db      5431 GAGGCTAATAAGAGTCGATACCAGTCTCGCAAAGTGGAGAACATGAAAGCCACGGTGGTG 5490

Qy       834 GAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGAATCCTGTCGAGTGT 893
Db      5491 GACAGGCTCACATCGGGGGCCAGATTGTACACGGGAGCGGACGTAGGCCGCATACCAACA 5550

Qy       894 TACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTAACAGGTTTACATCT 953
Db      5551 TACGCGGTTCGGTACCCCCGCCCCGTGTACTCCCCTACCGTGATCGAAAGATTCTCAAGC 5610

Qy       954 GCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATTACCCTACAGTAGCC 1013
Db      5611 CCCGATGTAGCAATCGCAGCGTGCAACGAATACCTATCCAGAAATTACCCAACAGTGGCG 5670

Qy      1014 AGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGGACGGCGCATCGTGC 1073
Db      5671 TCGTACCAGATAACAGATGAATACGACGCATACTTGGACATGGTTGACGGGTCGGATAGT 5730

Qy      1074 TGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACCCAAAGAAGCATAGC 1133
Db      5731 TGCTTGGACAGAGCGACATTCTGCCCGGCGAAGCTCCGGTGCTACCCGAAACATCATGCG 5790

Qy      1134 TATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGAATACATTACAAAAT 1193
Db      5791 TACCACCAGCCGACTGTACGCAGTGCCGTCCCGTCACCCTTTCAGAACACACTACAGAAC 5850

Qy      1194 GTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGCGAGAATTACCTGTC 1253
Db      5851 GTGCTAGCGGCCGCCACCAAGAGAAACTGCAACGTCACGCAAATGCGAGAACTACCCACC 5910

Qy      1254 TTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCATGCAATGATGAGTAC 1313
Db      5911 ATGGACTCGGCAGTGTTCAACGTGGAGTGCTTCAAGCGCTATGCCTGCTCCGGAGAATAT 5970

Qy      1314 TGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACGTTACGCAATATGTG 1373
Db      5971 TGGGAAGAATATGCTAAACAACCTATCCGGATAACCACTGAGAACATCACTACCTATGTG 6030

Qy      1374 ACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTCATAATCTAAAACCG 1433
Db      6031 ACCAAATTGAAAGGCCCGAAAGCTGCTGCCTTGTTCGCTAAGACCCACAACTTGGTTCCG 6090

Qy      1434 TTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAGATGTCAAAGTTACT 1493
Db      6091 CTGCAGGAGGTTCCCATGGACAGATTCACGGTCGACATGAAACGAGATGTCAAAGTCACT 6150

Qy      1494 CCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTCAGGCTGCAGATCCC 1553
Db      6151 CCAGGGACGAAACACACAGAGGAAAGACCCAAAGTCCAGGTAATTCAAGCAGCGGAGCCA 6210

Qy      1554 CTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTAGACTGAATGCGGTG 1613
Db      6211 TTGGCGACCGCTTACCTGTGCGGCATCCACAGGGAATTAGTAAGGAGACTAAATGCTGTG 6270

Qy      1614 CTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATTTTGATGCGATTATT 1673
Db      6271 TTACGCCCTAACGTGCACACATTGTTTGATATGTCGGCCGAAGACTTTGACGCGATCATC 6330

Qy      1674 GCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCGCGTCGTTTGATAAA 1733
Db      6331 GCCTCTCACTTCCACCCAGGAGACCCGGTTCTAGAGACGGACATTGCATCATTCGACAAA 6390

Qy      1734 AGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGGACTTAGGTGTCGAC 1793
Db      6391 AGCCAGGACGACTCCTTGGCTCTTACAGGTTTAATGATCCTCGAAGATCTAGGGGTGGAT 6450

Qy      1794 CAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACATCTGTGCACCTACCT 1853
Db      6451 CAGTACCTGCTGGACTTGATCGAGGCAGCCTTTGGGGAAATATCCAGCTGTCACCTACCA 6510

Qy      1854 ACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGTTCTTAACGCTGTTT 1913
Db      6511 ACTGGCACGCGCTTCAAGTTCGGAGCTATGATGAAATCGGGCATGTTTCTGACTTTGTTT 6570

Qy      1914 GTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTGAACGGTTAACCACG 1973
Db      6571 ATTAACACTGTTTTGAACATCACCATAGCAAGCAGGGTACTGGAGCAGAGACTCACTGAC 6630

Qy      1974 TCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTGTCGTCTCCGACACC 2033
Db      6631 TCCGCCTGTGCGGCCTTCATCGGCGACGACAACATCGTTCACGGAGTGATCTCCGACAAG 6690

Qy      2034 TTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAATTATTGATGCAGTT 2093
Db      6691 CTGATGGCGGAGAGGTGCGCGTCGTGGGTCAACATGGAGGTGAAGATCATTGACGCTGTC 6750

Qy      2094 ATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGGACCAGATAACAGGC 2153
Db      6751 ATGGGCGAAAAACCCCCATATTTTTGTGGGGGATTCATAGTTTTTGACAGCGTCACACAG 6810

Qy      2154 ACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTGGAAAACCATTGCCA 2213
Db      6811 ACCGCCTGCCGTGTTTCAGACCCACTTAAGCGCCTGTTCAAGTTGGGTAAGCCGCTAACA 6870

Qy      2214 GTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATGAAGCAATGCGATGG 2273
Db      6871 GCTGAAGACAAGCAGGACGAAGACAGGCGACGAGCACTGAGTGACGAGGTTAGCAAGTGG 6930

Qy      2274 AACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCAGATACGAGATCATA 2333
Db      6931 TTCCGGACAGGCTTGGGGGCCGAACTGGAGGTGGCACTAACATCTAGGTATGAGGTAGAG 6990

Qy      2334 CTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCGTTAAGAACTTCAAG 2393
Db      6991 GGCTGCAAAAGTATCCTCATAGCCATGGCCACCTTGGCGAGGGACATTAAGGCGTTTAAG 7050

Qy      2394 AGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGGTGACGTAGTAGACA 2453
Db      7051 AAATTGAGAGGACCTGTTATACACCTCTACGGCGGTCCTAGATTGGTGCGTTAATACACA 7110
```

RESULT 15
US-09-994-412-3
; Sequence 3, Application US/09994412
; Publication No. US20030059943A1
; GENERAL INFORMATION:
; APPLICANT: Certa, Ulrich
; APPLICANT: Lundstrom, Kenneth
; TITLE OF INVENTION: Inhibition of Expression of a Target Gene
; FILE REFERENCE: 20787
; CURRENT APPLICATION NUMBER: US/09/994,412
; CURRENT FILING DATE: 2002-03-12

```
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
;   LENGTH: 10610
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: pSFV2gen(PD) vector
US-09-994-412-3

  Query Match            11.8%;  Score 764;  DB 11;  Length 10610;
  Best Local Similarity   63.2%;  Pred. No. 8.8e-217;
  Matches 1175;  Conservative    0;  Mismatches  685;  Indels    0;  Gaps    0;

Qy       594 GCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTCAACAGAAATCAGTA 653
            |||||| || |||||||||||| ||||| || |||   || ||  | ||||| |||||| ||
Db      5529 GCGGGTGCATATATTTTCTCCTCGGACACTGGCAGCGGACATTTACAACAAAAATCCGTT 5588

Qy       654 CGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATGAGAAGTATTACGCC 713
             | ||     || || ||      |    | |||||  || |     | |||||    ||| |
Db      5589 AGGCAGCACAATCTCCAGTGCGCACAACTGGATGCGGTCCAGGAGGAGAAAATGTACCCG 5648

Qy       714 CCGCGCCTCGATCTCGAAAAGAGAGAAAATGTTACAGAAGAAACTGCAATTATGCGCCTCT 773
             ||     | |||   || || |||||  |||| | |  |||| ||||   |   | | ||
Db      5649 CCAAAATTGGATACTGAGAGGGAGAAGCTGTTGCTGCTGAAAATGCAGATGCACCCATCG 5708

Qy       774 GAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGAAAGCAATTACAGCG 833
             || |  ||||  ||  | || || || || ||||| || || |||||||| |      | |
Db      5709 GAGGCTAATAAGAGTCGATACCAGTCTCGCAAAGTGGAGAACATGAAAGCCACGGTGGTG 5768

Qy       834 GAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGAATCCTGTCGAGTGT 893
             ||  | ||||  || ||      |     ||       | | || ||    |   |
Db      5769 GACAGGCTCACATCGGGGGCCAGATTGTACACGGGAGCGGACGTAGGCCGCATACCAACA 5828

Qy       894 TACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTAACAGGTTTACATCT 953
             |||   ||    || ||    ||  | |||||  | || || ||  | || ||  ||
Db      5829 TACGCGGTTCGGTACCCCCGCCCCGTGTACTCCCCTACCGTGATCGAAAGATTCTCAAGC 5888

Qy       954 GCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATTACCCTACAGTAGCC 1013
             | || || ||  |   | ||||||||  |    |       |||||||| ||||| ||
Db      5889 CCCGATGTAGCAATCGCAGCGTGCAACGAATACCTATCCAGAAATTACCCAACAGTGGCG 5948

Qy      1014 AGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGGACGGCGCATCGTGC 1073
               ||    |||||||||||||||| || ||  | ||||||| |||||  |      |
Db      5949 TCGTACCAGATAACAGATGAATACGACGCATACTTGGACATGGTTGACGGGTCGGATAGT 6008

Qy      1074 TGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACCCAAAGAAGCATAGC 1133
             ||  | || | ||| || || || ||||| || ||  |  ||||||| ||  | |||
Db      6009 TGCTTGGACAGAGCGACATTCTGCCCGGCGAAGCTCCGGTGCTACCCGAAACATCATGCG 6068

Qy      1134 TATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGAATACATTACAAAAT 1193
             ||    |||||     || |    |||||||| || ||  | ||||| ||| |||| ||
Db      6069 TACCACCAGCCGACTGTACGCAGTGCCGTCCCGTCACCCTTTCAGAACACACTACAGAAC 6128

Qy      1194 GTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGCGAGAATTACCTGTC 1253
             ||  | || || || || || || ||| ||||||| || |||||||||||| ||||   |
Db      6129 GTGCTAGCGGCTGCCACCAAGAGAAACTGCAACGTCACGCAAATGCGAGAACTACCCACC 6188

Qy      1254 TTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCATGCAATGATGAGTAC 1313
               | || ||||| |  || || || || || ||||||   || || |||   |  || ||
Db      6189 ATGGACTCGGCAGTGTTCAACGTGGAGTGCTTCAAGCGCTATGCCTGCTCCGGAGAATAT 6248

Qy      1314 TGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACGTTACGCAATATGTG 1373
             |||||     | |     |  | |||||| ||| |||| || |||||| | ||    ||||||
Db      6249 TGGGAAGAATATGCTAAACAACCTATCCGGATAACCACTGAGAACATCACTACCTATGTG 6308

Qy      1374 ACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTCATAATCTAAAACCG 1433
             || ||  |||||||  |||||||| || || |||||  || || || || ||  |    |||
Db      6309 ACCAAATTGAAAGGCCCGAAAGCTGCTGCCTTGTTCGCTAAGACCCACAACTTGGTTCCG 6368

Qy      1434 TTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAGATGTCAAAGTTACT 1493
              ||||||||  | || |||||    ||||     | ||   | ||  ||||||||||||| |||
Db      6369 CTGCAGGAGGTTCCCATGGACAGATTCACGGTCGACATGAAACGAGATGTCAAAGTCACT 6428

Qy      1494 CCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTCAGGCTGCAGATCCC 1553
             || || |||||||| ||||||||  | || || || ||||| ||||| || || || ||
Db      6429 CCAGGGACGAAACACACAGAGGAAAGACCCAAAGTCCAGGTAATTCAAGCAGCGGAGCCA 6488

Qy      1554 CTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTAGACTGAATGCGGTG 1613
               | || |||||||||| ||||| |||||  ||||||||||  | ||||| ||||| |||
Db      6489 TTGGCGACCGCTTACCTGTGCGGCATCCACAGGGAATTAGTAAGGAGACTAAATGCTGTG 6548

Qy      1614 CTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATTTTGATGCGATTATT 1673
               | |  || || |  || ||  | || || ||||| || ||||| ||||| |||||  ||
Db      6549 TTACGCCCTAACGTGCACACATTGTTTGATATGTCGGCCGAAGACTTTGACGCGATCATC 6608

Qy      1674 GCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCGCGTCGTTTGATAAA 1733
             ||     || |||||||  || ||||| ||  | || |||||||| || || || || |||
Db      6609 GCCTCTCACTTCCACCCAGGAGACCCGGTTCTAGAGACGGACATTGCATCATTCGACAAA 6668

Qy      1734 AGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGGACTTAGGTGTCGAC 1793
             ||| | ||||||  |  | ||  || || |  || |||||||| || ||  |||| || ||
Db      6669 AGCCAGGACGACTCCTTGGCTCTTACAGGTTTAATGATCCTCGAAGATCTAGGGGTGGAT 6728

Qy      1794 CAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACATCTGTGCACCTACCT 1853
             ||     ||  | || ||||| ||||| || || ||  | ||  |          ||||||||
Db      6729 CAGTACCTGCTGGACTTGATCGAGGCAGCCTTTGGGGAAATATCCAGCTGTCACCTACCA 6788

Qy      1854 ACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGTTCTTAACGCTGTTT 1913
             || || ||| | || || || || || |||||||||| || |||||  | ||  |||||
Db      6789 ACTGGCACGCGCTTCAAGTTCGGAGCTATGATGAAATCGGGCATGTTTCTGACTTTGTTT 6848

Qy      1914 GTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTGAACGGTTAACCACG 1973
               |||||  |  | || ||||  || || |||||  |||||      |  |  | ||
Db      6849 ATTAACACTGTTTTGAACATCACCATAGCAAGCAGGGTACTGGAGCAGAGACTCACTGAC 6908

Qy      1974 TCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTGTCGTCTCCGACACC 2033
             || || || ||||||||  |||||||||||  |||||  || || || ||  |||||||||
Db      6909 TCCGCCTGTGCGGCCTTCATCGGCGACGACAACATCGTTCACGGAGTGATCTCCGACAAG 6968

Qy      2034 TTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAATTATTGATGCAGTT 2093
             |||||||||||||| |||||  | ||| | |||||||| || || || |||| || ||
Db      6969 CTGATGGCGGAGAGGTGCGCGTCGTGGGTCAACATGGAGGTGAAGATCATTGACGCTGTC 7028

Qy      2094 ATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGGACCAGATAACAGGC 2153
             || ||   ||| | || || || |||||||||| ||  |  | |||    | |||
Db      7029 ATGGGCGAAAAACCCCCATATTTTTGTGGGGGATTCATAGTTTTTGACAGCGTCACACAG 7088

Qy      2154 ACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTGGAAAACCATTGCCA 2213
             || |||||| | ||  |||||||| || ||  | || || ||| | || || || |  ||
Db      7089 ACCGCCTGCCGTGTTTCAGACCCACTTAAGCGCCTGTTCAAGTTGGGTAAGCCGCTAACA 7148

Qy      2214 GTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATGAAGCAATGCGATGG 2273
             |  || || |  || |||   ||| | || || ||||||  ||| || |  |     |||
Db      7149 GCTGAAGACAAGCAGGACGAAGACAGGCGACGAGCACTGAGTGACGAGGTTAGCAAGTGG 7208

Qy      2274 AACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCAGATACGAGATCATA 2333
               | | |  ||  |   || ||| | | |  |||  ||  ||| || || ||| |
Db      7209 TTCCGGACAGGCTTGGGGGCCGAACTGGAGGTGGCACTAACATCTAGGTATGAGGTAGAG 7268

Qy      2334 CTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCGTTAAGAACTTCAAG 2393
                     ||| |||   |  || |||| || ||      | |||||   || |||
Db      7269 GGCTGCAAAAGTATCCTCATAGCCATGGCCACCTTGGCGAGGGACATTAAGGCGTTTAAG 7328

Qy      2394 AGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGGTGACGTAGTAGACA 2453
             |    | |||||         ||   |||||||||| | |||| || ||||    || || |||
Db      7329 AAATTGAGAGGACCTGTTATACACCTCTACGGCGGTCCTAGATTGGTGCGTTAATACACA 7388

Search completed: November 16, 2003, 14:53:10
Job time : 1807 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 22:40:28 ; Search time 326 Seconds
(without alignments)
8780.275 Million cell updates/sec

Title: US-10-023-649-1_COPY_5000_11484
Perfect score: 6485
Sequence: 1 caccgccagcaacgacaacc.....tattttgtttttaaaatttc 6485

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

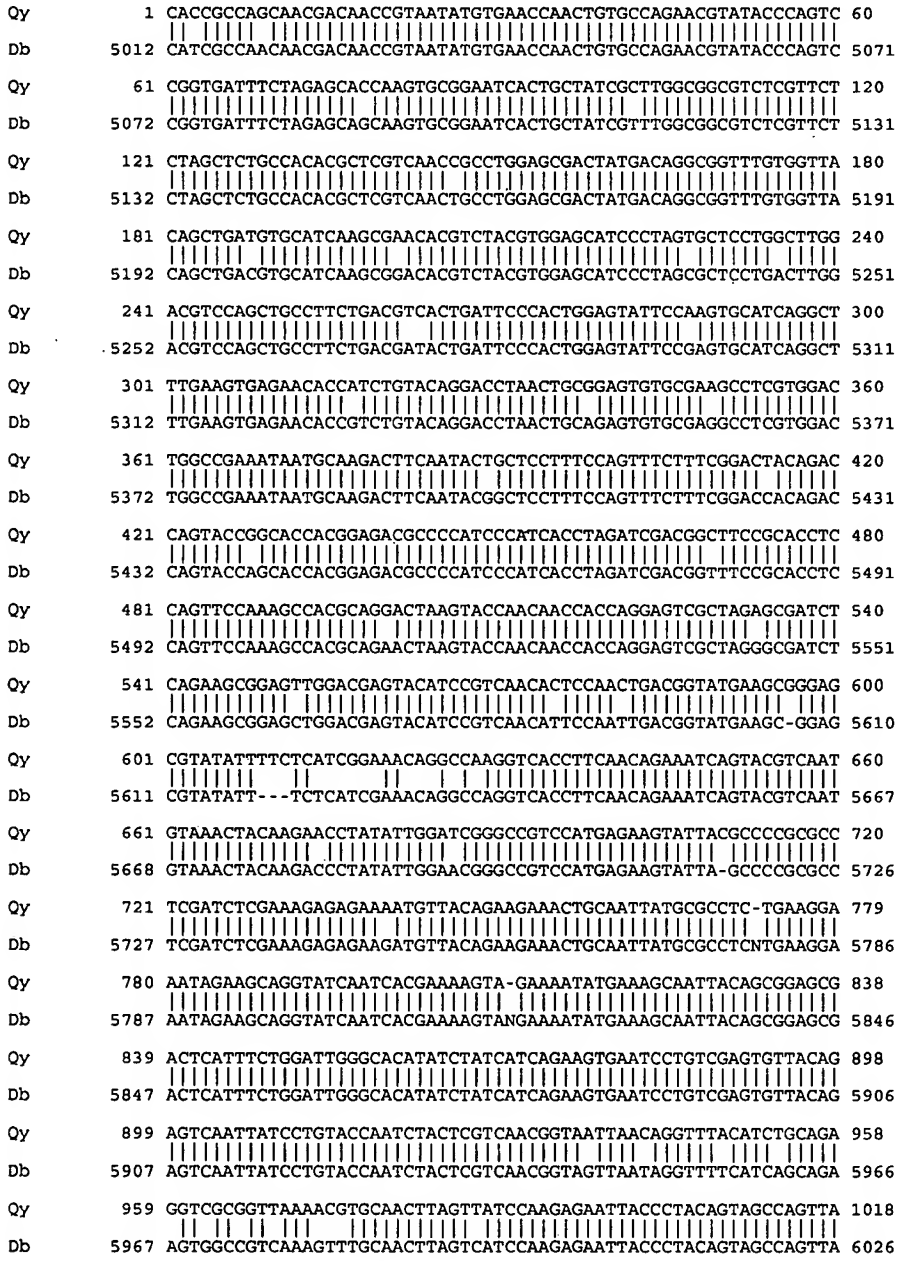
SUMMARIES

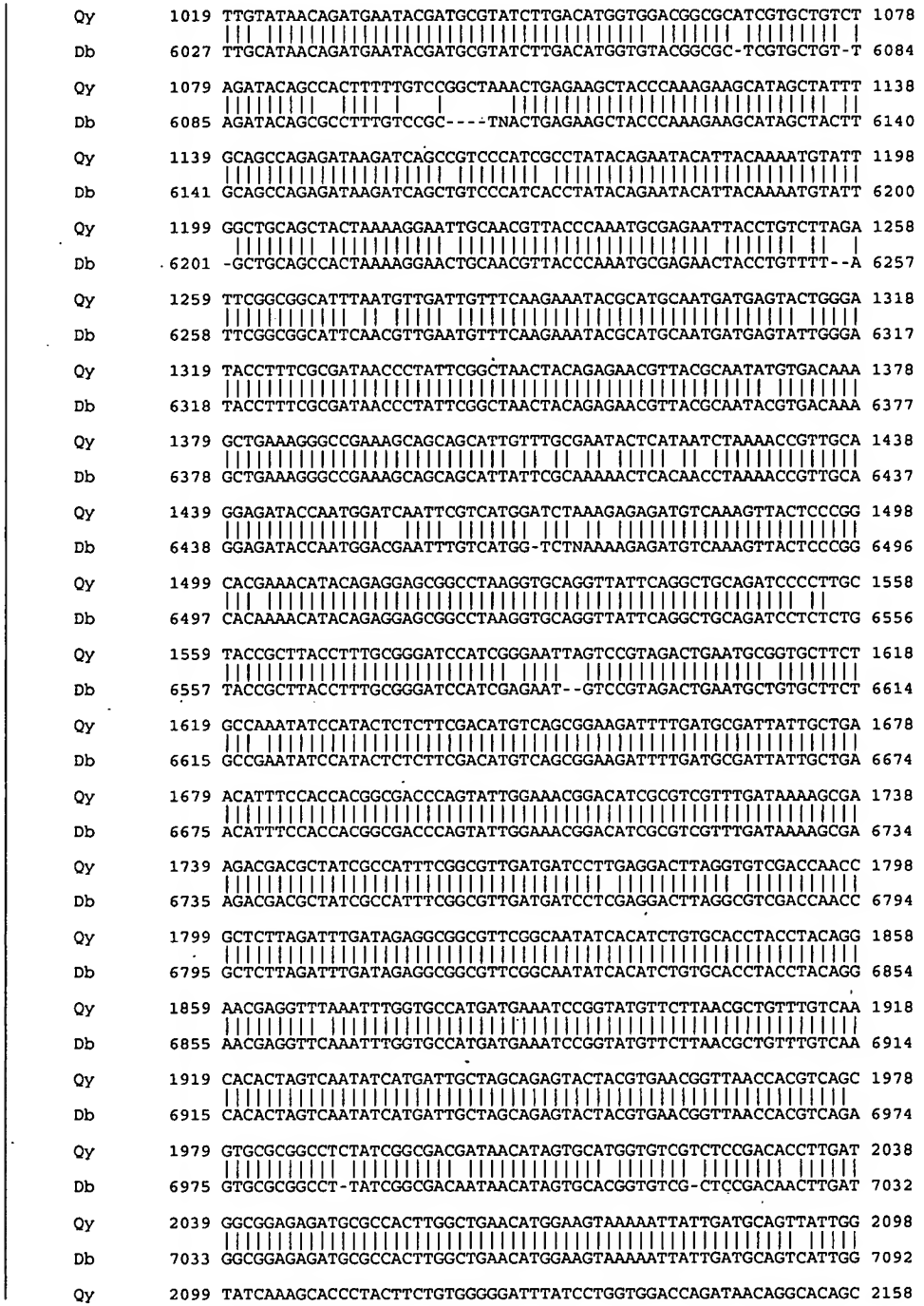
Result No.	Score	% Query Match	Length	DB	ID	Description
1	5879.2	90.7	11492	3	US-08-991-840A-1	Sequence 1, Appli
2	2397.6	37.0	11663	1	US-08-801-263A-7	Sequence 7, Appli
3	2397.6	37.0	11663	3	US-09-102-248-7	Sequence 7, Appli
4	2397.6	37.0	11663	4	US-09-367-764-7	Sequence 7, Appli
5	2396	36.9	11663	1	US-08-446-932-1	Sequence 1, Appli
6	2396	36.9	11663	1	US-08-801-263A-1	Sequence 1, Appli
7	2396	36.9	11663	3	US-09-102-248-1	Sequence 1, Appli
8	2396	36.9	11663	4	US-09-367-764-1	Sequence 1, Appli
9	2388.4	36.8	11703	1	US-08-801-263A-8	Sequence 8, Appli
10	2388.4	36.8	11703	3	US-09-102-248-8	Sequence 8, Appli
11	2388.4	36.8	11703	4	US-09-367-764-8	Sequence 8, Appli
12	2386.4	36.8	11717	1	US-08-801-263A-4	Sequence 4, Appli
13	2386.4	36.8	11717	3	US-09-102-248-4	Sequence 4, Appli
14	2386.4	36.8	11717	4	US-09-367-764-4	Sequence 4, Appli
15	2380.4	36.7	11740	4	US-09-415-784-103	Sequence 103, App
16	2380.4	36.7	11740	4	US-09-415-785A-103	Sequence 103, App
17	2380.4	36.7	11740	4	US-08-944-465-103	Sequence 103, App
18	2380.4	36.7	11740	4	US-09-415-868-103	Sequence 103, App
19	2380.4	36.7	11740	4	US-09-415-900-103	Sequence 103, App
20	2369.2	36.5	16656	1	US-08-741-881-1	Sequence 1, Appli
21	2369.2	36.5	16656	1	US-08-739-158-1	Sequence 1, Appli
22	2369.2	36.5	16656	2	US-08-739-167-1	Sequence 1, Appli
23	2369.2	36.5	16656	3	US-08-404-796-1	Sequence 1, Appli
24	2369.2	36.5	16656	3	US-08-931-869-1	Sequence 1, Appli
25	2369.2	36.5	16656	4	US-09-350-399-1	Sequence 1, Appli
26	2369.2	36.5	16656	4	US-09-236-140A-1	Sequence 1, Appli
27	2122.6	32.7	13905	3	US-08-972-218-1	Sequence 1, Appli
28	2122.6	32.7	13905	4	US-09-193-707-1	Sequence 1, Appli
29	1738.8	26.8	11464	3	US-08-991-840A-2	Sequence 2, Appli
30	1717.4	26.5	11459	3	US-09-454-721A-3	Sequence 3, Appli
31	1650.8	25.5	7970	4	US-09-193-707-6	Sequence 6, Appli
32	1646.6	25.4	9521	3	US-08-972-218-2	Sequence 2, Appli
33	1646.6	25.4	9521	4	US-09-193-707-2	Sequence 2, Appli
34	1442.6	22.2	11517	1	US-07-920-281C-1	Sequence 1, Appli
35	1442.6	22.2	11517	3	US-08-466-277-1	Sequence 1, Appli
36	863.2	13.3	4003	3	US-08-991-840A-3	Sequence 3, Appli
37	827.6	12.8	8000	4	US-09-415-784-102	Sequence 102, App
38	827.6	12.8	8000	4	US-09-415-785A-102	Sequence 102, App
39	827.6	12.8	8000	4	US-08-944-465-102	Sequence 102, App
40	827.6	12.8	8000	4	US-09-415-868-102	Sequence 102, App
41	827.6	12.8	8000	4	US-09-415-900-102	Sequence 102, App
42	824.4	12.7	8000	4	US-09-415-784-101	Sequence 101, App
43	824.4	12.7	8000	4	US-09-415-785A-101	Sequence 101, App
44	824.4	12.7	8000	4	US-08-944-465-101	Sequence 101, App
45	824.4	12.7	8000	4	US-09-415-868-101	Sequence 101, App

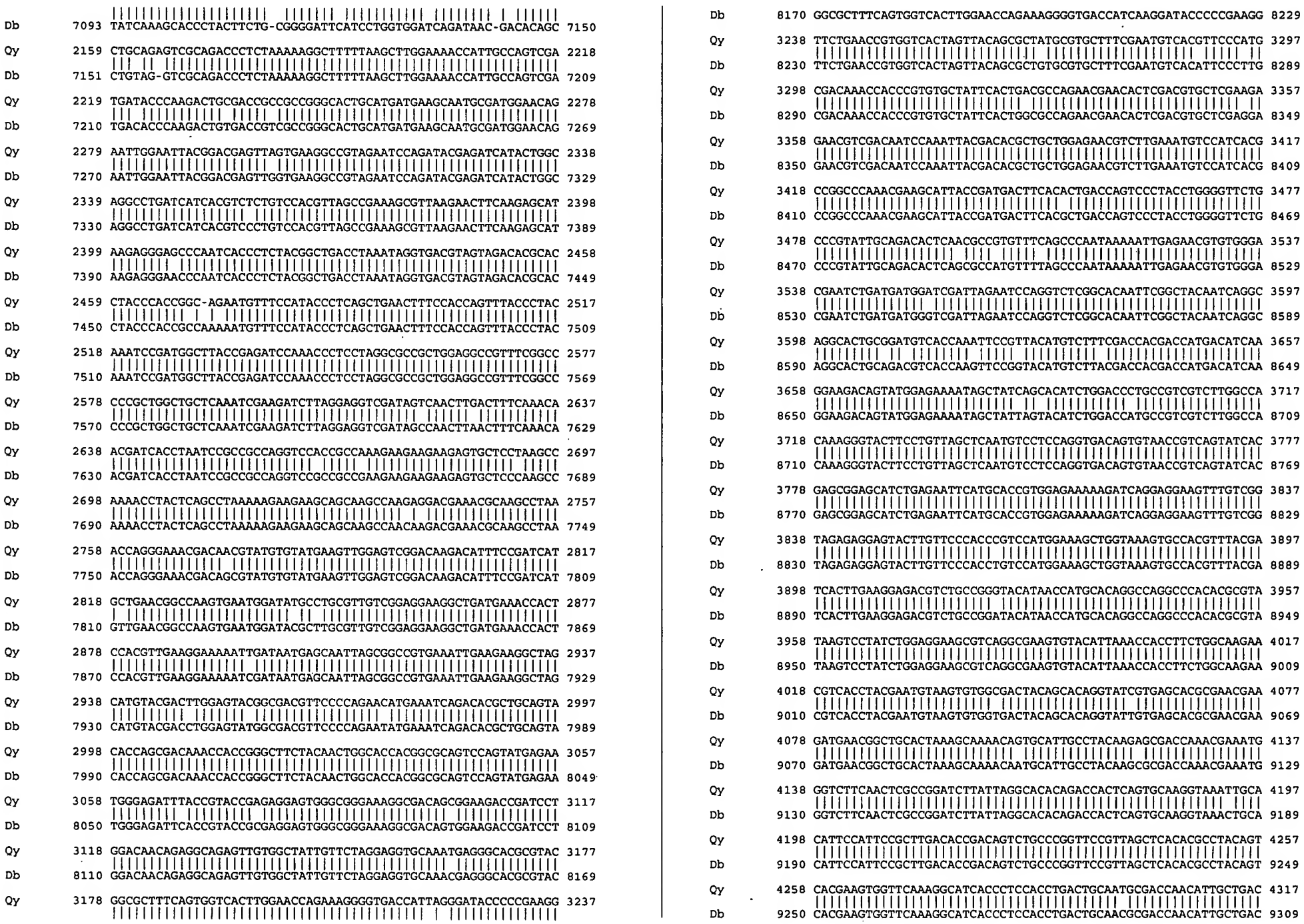
ALIGNMENTS

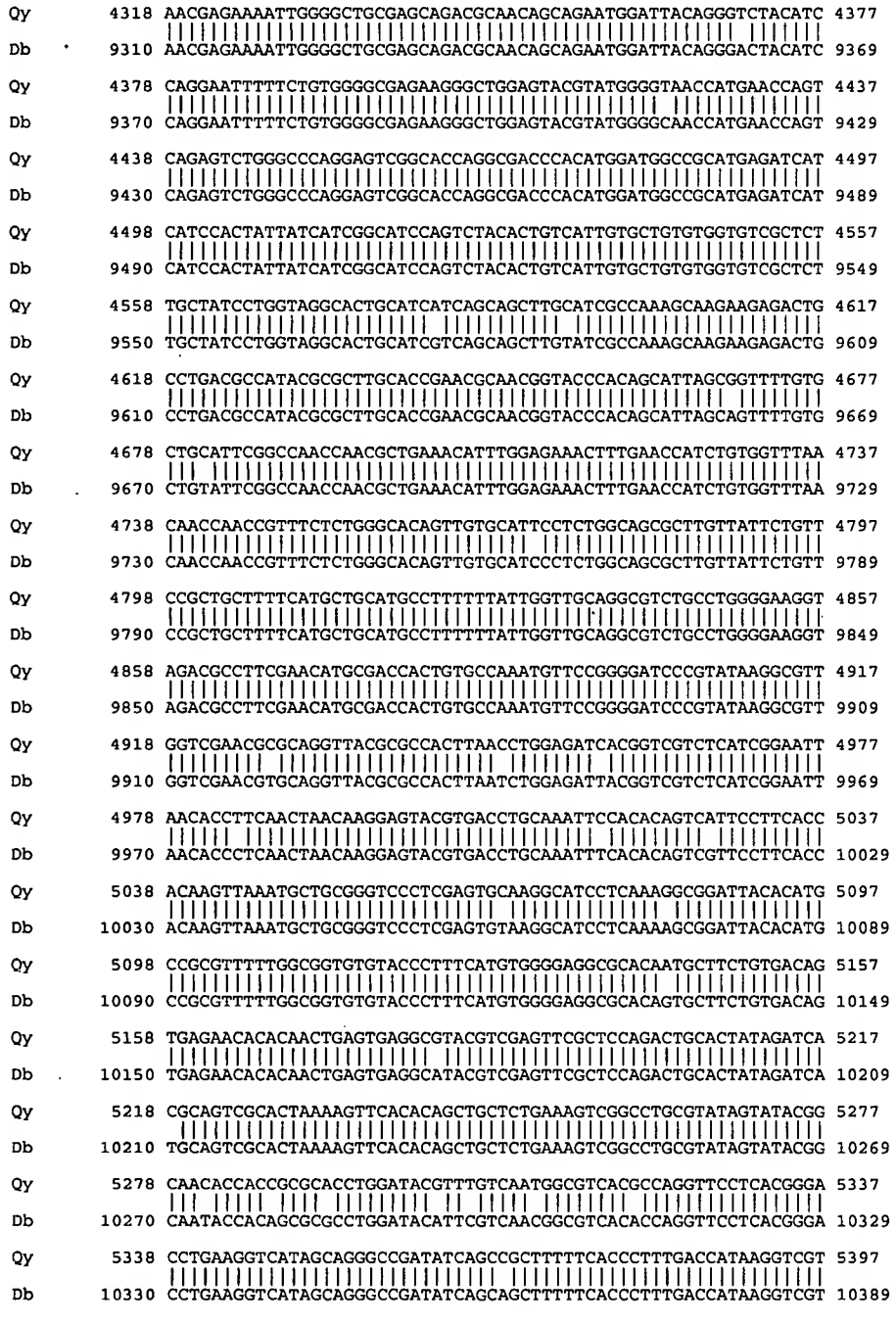
```
RESULT 1
US-08-991-840A-1
; Sequence 1, Application US/08991840A
; Patent No. 6261570
;  GENERAL INFORMATION:
;    APPLICANT:  Michael D. Parker
;    APPLICANT:  Jonathan F. Smith
;    APPLICANT:  Bruce Crise
;    APPLICANT:  Mark Steve Oberste
;    APPLICANT:  Shannon Schmura
;    TITLE OF INVENTION:  Live Attenuated Virus Vaccines for Eastern Equine Encephalitis
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Charles H. Harris
;      STREET:  USA MRMC -  MRMC-JA
;      CITY:  FORT DETRICK, FREDERICK
;      STATE:  MARYLAND
;      COUNTRY:  USA
;      ZIP:  21702-5012
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh 7.5
;      SOFTWARE:  Microsoft Word 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/991,840A
;      FILING DATE:  December 16, 1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  Provisional Application 60/047162,
;      FILING DATE:  May 20, 1997
;      APPLICATION NUMBER:  Provisional Application 60/053,652
;      FILING DATE:  July 24, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles H. Harris
;      REGISTRATION NUMBER:  34,616
;      REFERENCE/DOCKET NUMBER:  003/058/SAP RIID 96-01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 619-2065
;      TELEFAX:  (301) 619-5034
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11492 base pairs
;      TYPE:  Nucleic acid
;      STRANDEDNESS:  Double
;      TOPOLOGY:  Linear
;    FEATURE:  OTHER INFORMATION:  N at all occurences is = unknown.
;    FEATURE:  OTHER INFORMATION: K at all
;    FEATURE:  occurences is = G or T
US-08-991-840A-1
```

Query Match 90.7%; Score 5879.2; DB 3; Length 11492;
Best Local Similarity 96.2%; Pred. No. 0;
Matches 6239; Conservative 0; Mismatches 220; Indels 29; Gaps 20;

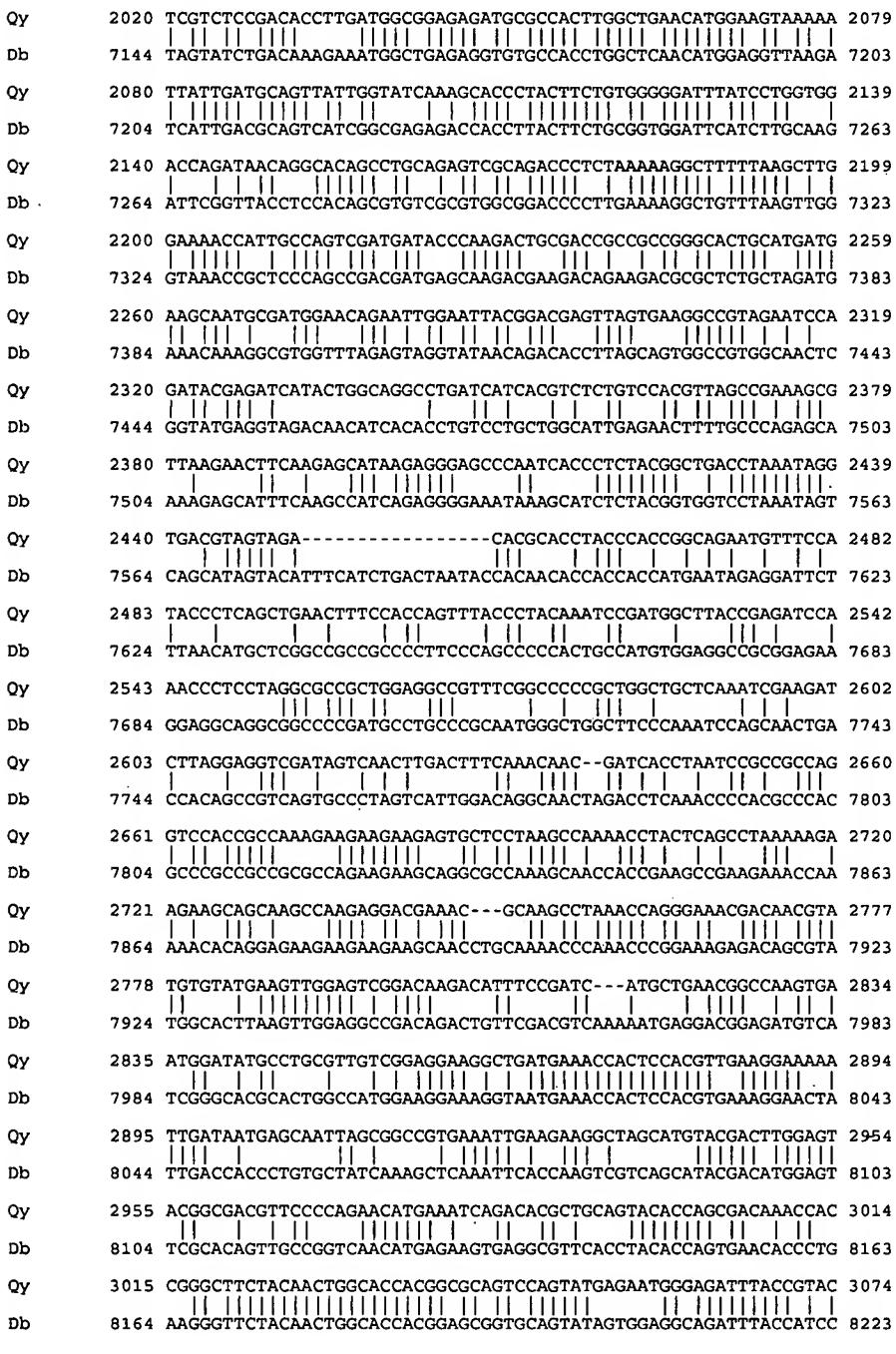


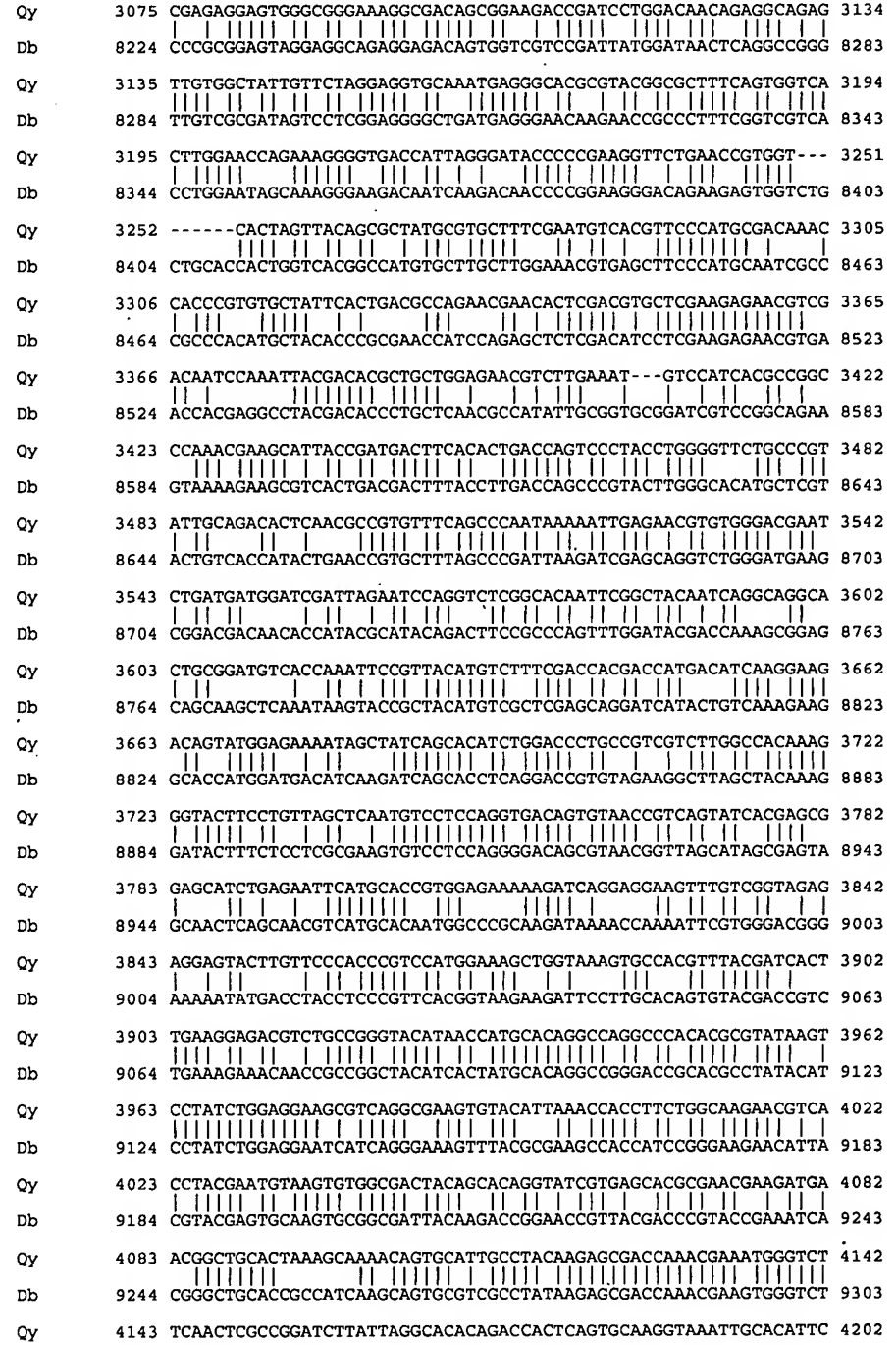


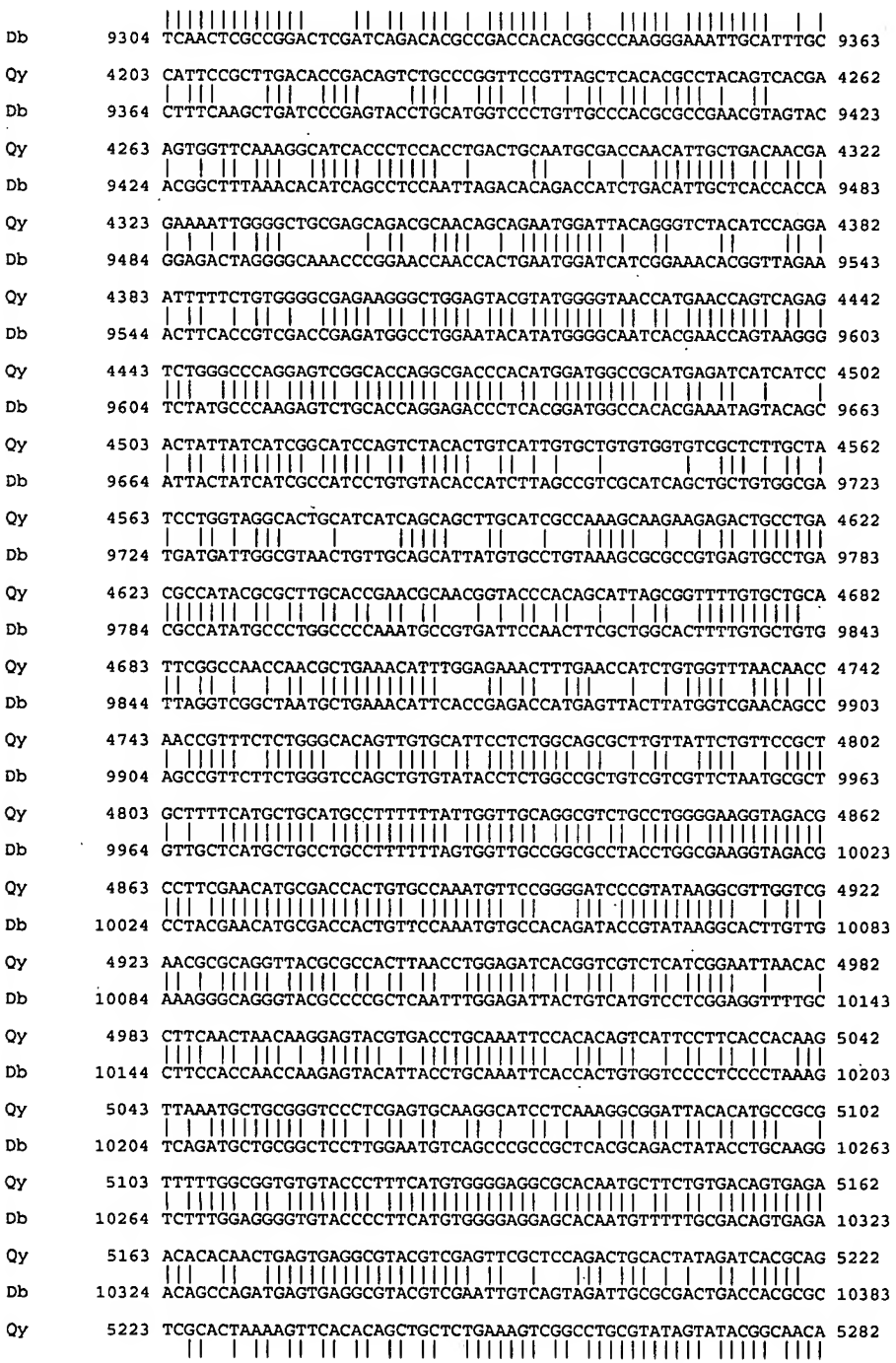



Qy 4318 AACGAGAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATC 4377
Db 9310 AACGAGAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGACTACATC 9369
Qy 4378 CAGGAATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGT 4437
Db 9370 CAGGAATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGCAACCATGAACCAGT 9429
Qy 4438 CAGAGTCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCAT 4497
Db 9430 CAGAGTCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCAT 9489
Qy 4498 CATCCACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCT 4557
Db 9490 CATCCACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCT 9549
Qy 4558 TGCTATCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTG 4617
Db 9550 TGCTATCCTGGTAGGCACTGCATCGTCAGCAGCTTGTATCGCCAAAGCAAGAAGAGACTG 9609
Qy 4618 CCTGACGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTG 4677
Db 9610 CCTGACGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCAGTTTTGTG 9669
Qy 4678 CTGCATTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAA 4737
Db 9670 CTGTATTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAA 9729
Qy 4738 CAACCAACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTT 4797
Db 9730 CAACCAACCGTTTCTCTGGGCACAGTTGTGCATCCCTCTGGCAGCGCTTGTTATTCTGTT 9789
Qy 4798 CCGCTGCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGT 4857
Db 9790 CCGCTGCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGT 9849
Qy 4858 AGACGCCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTT 4917
Db 9850 AGACGCCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTT 9909
Qy 4918 GGTCGAACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATT 4977
Db 9910 GGTCGAACGTGCAGGTTACGCGCCACTTAATCTGGAGATTACGGTCGTCTCATCGGAATT 9969
Qy 4978 AACACCTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACC 5037
Db 9970 AACACCCTCAACTAACAAGGAGTACGTGACCTGCAAATTTCACACAGTCGTTCCTTCACC 10029
Qy 5038 ACAAGTTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATG 5097
Db 10030 ACAAGTTAAATGCTGCGGGTCCCTCGAGTGTAAGGCATCCTCAAAAGCGGATTACACATG 10089
Qy 5098 CCGCGTTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAG 5157
Db 10090 CCGCGTTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAGTGCTTCTGTGACAG 10149
Qy 5158 TGAGAACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCA 5217
Db 10150 TGAGAACACACAACTGAGTGAGGCATACGTCGAGTTCGCTCCAGACTGCACTATAGATCA 10209
Qy 5218 CGCAGTCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGG 5277
Db 10210 TGCAGTCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGG 10269
Qy 5278 CAACACCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGA 5337
Db 10270 CAATACCACAGCGCGCCTGGATACATTCGTCAACGGCGTCACACCAGGTTCCTCACGGGA 10329
Qy 5338 CCTGAAGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGT 5397
Db 10330 CCTGAAGGTCATAGCAGGGCCGATATCAGCAGCTTTTTCACCCTTTGACCATAAGGTCGT 10389
Qy 5398 CATCAGAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGG 5457
Db 10390 CATTAGAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAACCCAGG 10449
Qy 5458 AGCGTTCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGA 5517
Db 10450 AGCGTTCGGCGATATTCAAGCATCCTCTCTTGATGCCACAGACATAGTAGCCCGCACCGA 10509
Qy 5518 CATACGGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATC 5577
Db 10510 CATACGGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATC 10569
Qy 5578 AGGGTATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGG 5637
Db 10570 AGGGTATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTCGG 10629
Qy 5638 ATGTAAAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTAT 5697
Db 10630 ATGTAAAATTGAAGTGGAGCCTCTGCGAGCGACTAACTGTGCTTATGGGCACATCCCTAT 10689
Qy 5698 CTCGATTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGA 5757
Db 10690 CTCGATTGACATCCCTGATGCAGCTTTTGTGAGATCATCTGAATCACCAACAATTTTAGA 10749
Qy 5758 AGTTAGCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATT 5817
Db 10750 AGTCAGCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCGCTAACACT 10809
Qy 5818 ACAGTACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGT 5877
Db 10810 ACAGTACAAAGCTAACAGAGAGGGACATTGTCCAGTTCACTCCCACTCCACTACAGCTGT 10869
Qy 5878 TTTGAAGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCAC 5937
Db 10870 TTTGAAGGAAGCGACCACACATGTGACTGCCACAG---CCATAACACTACATTTTAGCAC 10926
Qy 5938 ATCGAGCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGC 5997
Db 10927 ATCGAGCCCACAAGCAAATTTCATAGTTTCGCTATGGCGCAAGAAGACCACCTGCAATGC 10986
Qy 5998 TGAATGTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATT 6057
Db 10987 TGAATGTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAGGTCGACCAAGAATT 11046
Qy 6058 CCAGGCGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATC 6117
Db 11047 CCAGGCGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATC 11106
Qy 6118 ATCCCTCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACG 6177
Db 11107 ATCCCTCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACG 11166
Qy 6178 TAGATGACTGAGCGCGGACACTGACATAGCGGT-AAAACTCGATGTACTTCCGAGGAAGC 6236
Db 11167 TAGATGACTGAGCGCGGACACTGACATAGCGGTAAAAACTCGATGTACTTCCGAGGAAGC 11226
Qy 6237 GTGGTGCATAATGCCACGCGCCGCTTGACACTAAAACTCGATGTATTTCCGAGGAAGCAC 6296
Db 11227 GTGGTGCATAATGCCACGCGCCGCTTGACACTAAAACTCGATGTATTTCCGAGGAAGCAC 11286
Qy 6297 AGTGCATAATGCTGTGCAGTGTCACATTAATCGTATATCACACTACATATTAACAACACT 6356
Db 11287 AGTGCATAATGCTGTGCAGTGTCACATTAATCGTATATTACACTACATATTAACAACACT 11346
Qy 6357 ATATCACTTTTATGAGACTCACTATGGGTCTCTAATATACACTACACATATTTTACTTAA 6416
Db 11347 ATATCACTTTTATGAGACTCACTATGGGTTTCTAATACACACTACACACATTTTATTTAA 11406
Qy 6417 AAACACTATACACACTTTATAAATTCTTTTATAATTTTTCTTTTGTTTTTATTTTGTTTT 6476
Db 11407 AAACACTACACACACTTTATAAATTCTTTTATAATTTTTCTTTTGTTTTTATTTTGTTT 11466
Qy 6477 TAAAATTT 6484

```
                | ||| ||
Db      11467 TTAAAATT 11474


RESULT 2
US-08-801-263A-7
; Sequence 7, Application US/08801263A
; Patent No. 5811407
;  GENERAL INFORMATION:
;    APPLICANT:  Johnston, Robert E.
;    APPLICANT:  Davis, Nancy L.
;    APPLICANT:  Simpson, Dennis A.
;    TITLE OF INVENTION:  System for the In Vivo Delivery and
;    TITLE OF INVENTION:  Expression of Heterologous Genes in the Bone Marrow
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson, P.A.
;      STREET:  1211 East Morehead Street
;      CITY:  Charlotte
;      STATE:  No. 5811407th Carolina
;      COUNTRY:  USA
;      ZIP:  28234
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/801,263A
;      FILING DATE:  19-FEB-1997
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sibley, Kenneth D.
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5470-147
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-420-2200
;      TELEFAX:  919-881-3175
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11663 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-801-263A-7

  Query Match             37.0%;  Score 2397.6;  DB 1;  Length 11663;
  Best Local Similarity   64.6%;  Pred. No. 0;
  Matches 3752;  Conservative    0;  Mismatches 1984;  Indels   74;  Gaps    9;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
              |||| |      |  |  || | ||| || || ||  |||| |||||||  || ||| | |
Db       5707 ACTGTCTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5766

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
              || |||| || || |  ||     | ||  |||||| |  ||||| ||    || | |
Db       5767 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTTCTGG 5826

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
              | |     |||||||||   |||||      | |||| ||  | |  |     |     |
Db       5827 AAAGAATCTACGCCCCGGTGCTCGACACGTCGAAAGAGGAACAGCTCAAACTCAGGTACC 5886

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
              |  |      || | ||||  || | |||||||||| || || ||||||||||||||   ||
Db       5887 AGATGATGCCCACCGAAGCCAACAAAAGCAGGTACCAGTCTCGAAAAGTAGAAAACCAGA 5946

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
              |||| || ||  | ||||||||  |||| ||  |  |    |||   ||  |   ||
Db       5947 AAGCCATAACCACTGAGCGACTGCTTTCAGGGCTACGGCTGTATAACTCTGCCACAG--- 6003

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
              |||     || || || |   |||  || ||   ||||   || ||    |  |||

Db       6004 ATCAGCCAGAATGCTATAAGATCACCTACCCGAAACCATCGTATTCCAGCAGTGTACCAG 6063

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
              |  |   |    || || | || ||       || |||     | | || |||||||
Db       6064 CGAACTACTCTGACCCAAAGTTGCTGTAGCTGTTTGTAACAACTATCTGCATGAGAATT 6123

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
              |||| || |||||   ||||   || || || || ||||||| ||  | || ||||| |
Db       6124 ACCCGACGGTAGCATCTTATCAGATCACCGACGAGTACGATGCTTACTTGGATATGGTAG 6183

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
              ||||  ||       || |||||||| || |||||||| || || || || ||||| ||||
Db       6184 ACGGGACAGTCGCTTGCCTAGATACTGCAACTTTTTGCCCCGCCAAGCTTAGAAGTTACC 6243

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
              | || |  ||    |||      ||| | ||  |    || || |||||  | || ||||
Db       6244 CGAAAAGACACGAGTATAGAGCCCCAAACATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6303

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
              | || || ||||| ||  |   ||| || |||||||| || |||||||| || |||||||
Db       6304 ACACGTTGCAAAACGTGCTCATTGCCGCGACTAAAAGAAACTGCAACGTCACACAAATGC 6363

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
              | ||| | ||     | || || ||| |||| || ||||| || ||    ||||| ||||
Db       6364 GTGAACTGCCAACACTGGACTCAGCGACATTCAACGTTGAATGCTTTCGAAAATATGCAT 6423

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
              ||||||| ||||| |||||    |||  |   || || ||| || | ||||| |||  ||
Db       6424 GCAATGACGAGTATTGGGAGGAGTTTGCCCGAAAGCCAATTAGGATCACTACTGAGTTCG 6483

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
              ||||   ||| ||| | |  |||||||| || || || || ||| |||| || || | |
Db       6484 TTACCGCATACGTGGCCAGACTGAAAGGCCCTAAGGCCGCCGCACTGTTCGCAAAGACGC 6543

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
              ||||| |    || ||||| ||  | || ||||||  |||||||||||  | || ||||
Db       6544 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6603

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
              | || |||||||| || |||||||||||| ||||| ||  | || || || || || || | |
Db       6604 ACGTGAAAGTTACACCTGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6663

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
              | || ||||| ||||| || |||||||||||| ||||||||||| ||||| |||| || |
Db       6664 AAGCCGCAGAACCCCTGGCGACCGCTTACCTATGCGGGATCCACCGGGAGTTAGTGCGCA 6723

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
              | || |  || ||  | || || || || || || ||||| ||||||| ||||| || |
Db       6724 GGCTTACAGCCGTTTTGCTACCCAACATTCACACGCTCTTTGACATGTCGGCGGAGGACT 6783

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
              ||||||| || || || ||||| ||| | || || ||||| ||| |||| ||||| ||||
Db       6784 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGTGACCCGGTACTGGAGACGGATATCG 6843

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
              | ||||| || |||||| |||||||||||| ||  |  | |  |||||||| | || |
Db       6844 CCTCGTTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGCCTGATGATCTTGGAAG 6903

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
              || | ||||| |||||||| ||  | || ||||| |||  || || ||  | ||  |||
Db       6904 ACCTGGGTGTGGACCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 6963

Qy       1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
              |    || || || || || ||  | || ||||| || || ||||||||||||||  ||||
Db       6964 CCACCCATCTGCCCACGGGTACCCGTTTCAAATTCGGGGCGATGATGAAATCCGGAATGT 7023

Qy       1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
              || | ||||| |||||||||||| |  | ||| || | || || |||||||| |    |
Db       7024 TCCTCACGCTCTTTGTCAACACAGTTCTGAATGTCGTTATCGCCAGCAGAGTATTGGAGG 7083

Qy       1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
              | ||| | |  ||||     || || || | |||||||||||| |||||  | || || |
Db       7084 AGCGGCTTAAAACGTCCAAATGTGCAGCATTTATCGGCGACGACAACATTATACACGGAG 7143
```

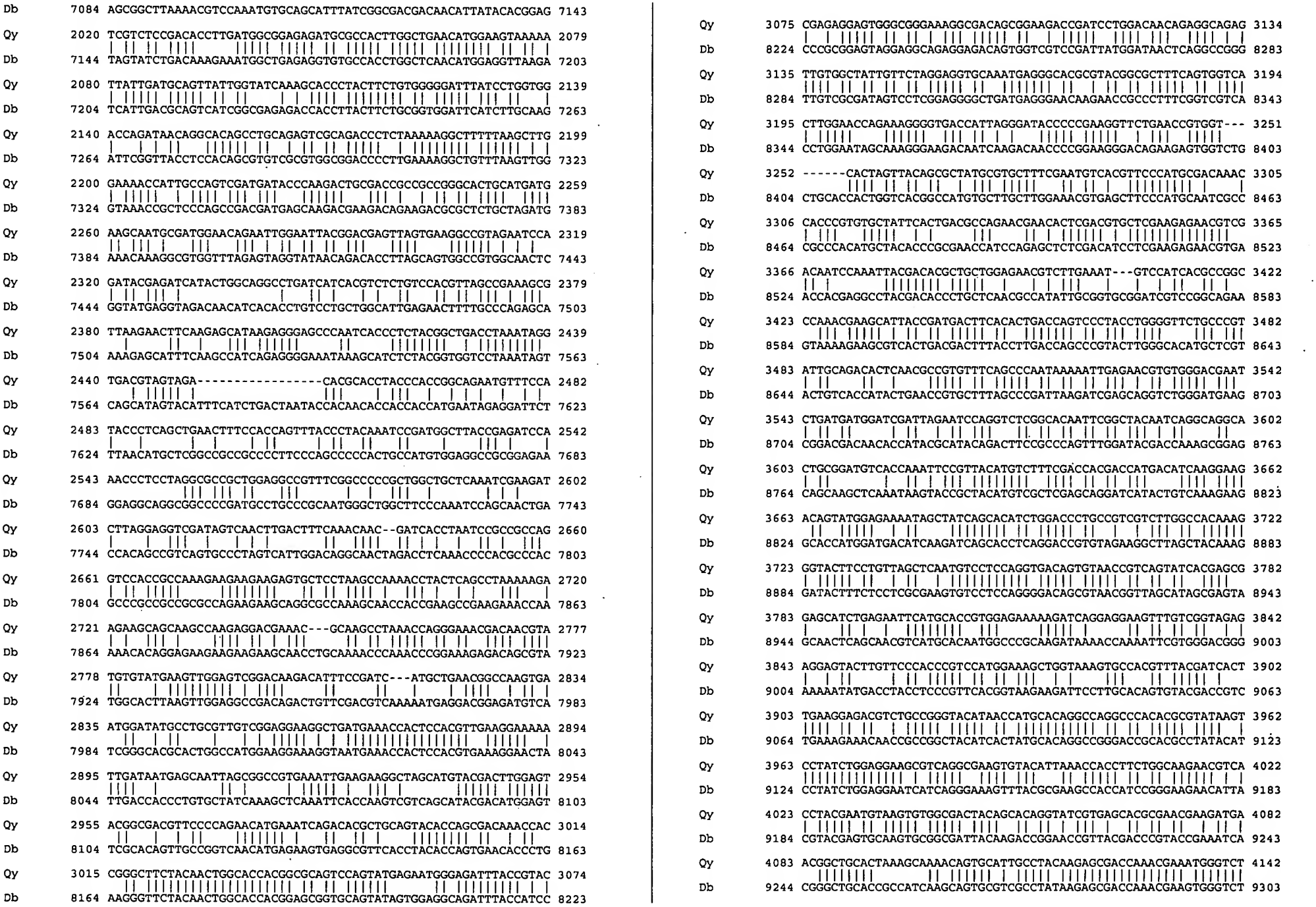


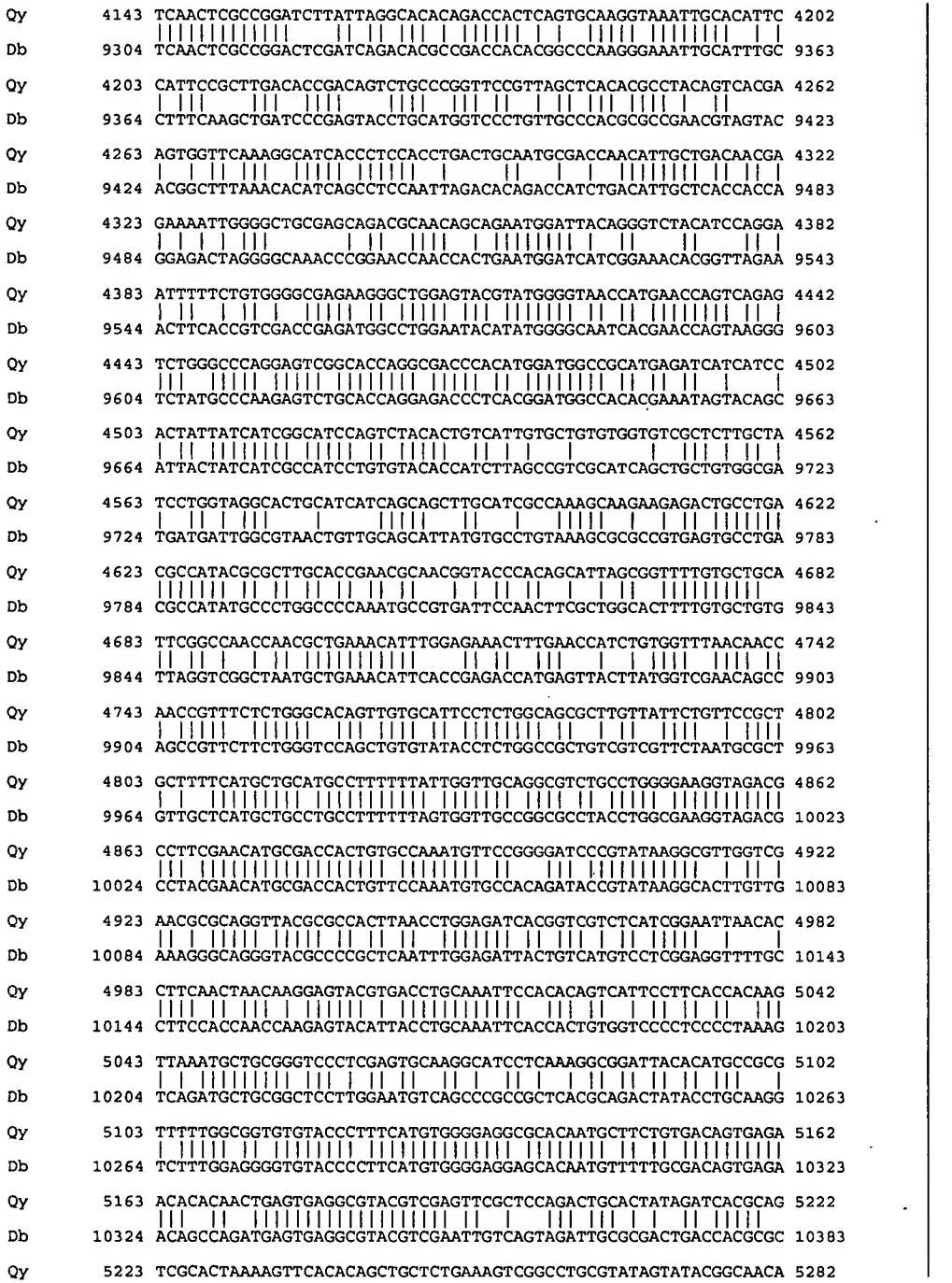



Db 9304 TCAACTCGCCGGACTCGATCAGACACGCCGACCACACGGCCCAAGGGAAATTGCATTTGC 9363
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9364 CTTTCAAGCTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAACGTAGTAC 9423
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9424 ACGGCTTTAAACACATCAGCCTCCAATTAGACACAGACCATCTGACATTGCTCACCACCA 9483
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9484 GGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCATCGGAAACACGGTTAGAA 9543
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9544 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGCAATCACGAACCAGTAAGGG 9603
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db 9604 TCTATGCCCAAGAGTCTGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9663
Qy 4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db 9664 ATTACTATCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTGCTGTGGCGA 9723
Qy 4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db 9724 TGATGATTGGCGTAACTGTTGCAGCATTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9783
Qy 4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db 9784 CGCCATATGCCCTGGCCCCAAATGCCGTGATTCCAACTTCGCTGGCACTTTTGTGCTGTG 9843
Qy 4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db 9844 TTAGGTCGGCTAATGCTGAAACATTCACCGAGACCATGAGTTACTTATGGTCGAACAGCC 9903
Qy 4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
Db 9904 AGCCGTTCTTCTGGGTCCAGCTGTGTATACCTCTGGCCGCTGTCGTCGTTCTAATGCGCT 9963
Qy 4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
Db 9964 GTTGCTCATGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10023
Qy 4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
Db 10024 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10083
Qy 4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db 10084 AAAGGGCAGGGTACGCCCCGCTCAATTTGGAGATTACTGTCATGTCCTCGGAGGTTTTGC 10143
Qy 4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db 10144 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCTAAAG 10203
Qy 5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db 10204 TCAGATGCTGCGGCTCCTTGGAATGTCAGCCCGCCGCTCACGCAGACTATACCTGCAAGG 10263
Qy 5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db 10264 TCTTTGGAGGGGTGTACCCCTTCATGTGGGGAGGAGCACAATGTTTTTGCGACAGTGAGA 10323
Qy 5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db 10324 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGTAGATTGCGCGACTGACCACGCGC 10383
Qy 5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db 10384 AGGCGATTAAGGTGCATACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10443
Qy 5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db 10444 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACCTGA 10503
Qy 5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db 10504 AAGTCATAGCTGGACCAATTTCAGCATTGTTTACACCATTCGATCACAAGGTCGTTATCA 10563
Qy 5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db 10564 ATCGCGGCCTGGTGTACAACTATGACTTTCCGGAATACGGAGCGATGAAACCAGGAGCGT 10623
Qy 5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db 10624 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAAGACCTCATCGCCAGCACAGACATTA 10683
Qy 5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db 10684 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCTGGAT 10743
Qy 5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db 10744 TCGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCCCCTTTTGGGTGCA 10803
Qy 5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db 10804 AGATTGCAGTCAATCCGCTTCGAGCGGTGGACTGCTCATACGGGAACATTCCCATTTCTA 10863
Qy 5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db 10864 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10923
Qy 5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db 10924 AATGTGATGTCAGTGAGTGCACTTATTCAGCGGACTTCGGAGGGATGGCTACCCTGCAGT 10983
Qy 5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db 10984 ATGTATCCGACCGCGAAGGACAATGCCCTGTACATTCGCATTCGAGCACAGCAACCCTCC 11043
Qy 5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db 11044 AAGAGTCGACAGTTCATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTCAGCACCGCGA 11103
Qy 5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db 11104 GCCCACAGGCGAACTTCATTGTATCGCTGTGTGGTAAGAAGACAACATGCAATGCAGAAT 11163
Qy 6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db 11164 GCAAACCACCAGCTGATCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTCCAAG 11223
Qy 6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db 11224 CCGCCATCTCAAAAACTTCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11283
Qy 6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
Db 11284 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11343
Qy 6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
Db 11344 GACCGCTACGCCCCAATGACCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11403
Qy 6242 GCATAATGCCACGCGCCGCT----------TGACACT 6268
Db 11404 GCATAATGCATCAGGCTGGTATATTAGATCCCCGCTTACCGCGGGCAATATAGCAACACC 11463
Qy 6269 AAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGT 6318
Db 11464 AAAACTCGACGTATTTCCGAGGAAGCGCAGTGCATAATGCTGCGCAGTGT 11513

```
RESULT 3
US-09-102-248-7
; Sequence 7, Application US/09102248
; Patent No. 6008035
;  GENERAL INFORMATION:
;    APPLICANT:  Johnston, Robert E.
;    APPLICANT:  Davis, Nancy L.
;    APPLICANT:  Simpson, Dennis A.
;    TITLE OF INVENTION:  System for the In Vivo Delivery and
;    TITLE OF INVENTION:  Expression of Heterologous Genes in the Bone Marrow
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson, P.A.
;      STREET:  1211 East Morehead Street
;      CITY:  Charlotte
;      STATE:  No. 6008035th Carolina
;      COUNTRY:  USA
;      ZIP:  28234
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/102,248
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/801,263
;      FILING DATE:  19-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sibley, Kenneth D.
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5470-147
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-420-2200
;      TELEFAX:  919-881-3175
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11663 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-102-248-7

  Query Match             37.0%;  Score 2397.6;  DB 3;  Length 11663;
  Best Local Similarity   64.6%;  Pred. No. 0;
  Matches 3752;  Conservative    0;  Mismatches 1984;  Indels   74;  Gaps    9;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
              |||| |     |  |  || | ||| || || ||  |||| |||||||  || ||| | |
Db       5707 ACTGTCTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5766

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
              || |||| || || |  ||     | ||   |||||| |  |||| ||    || |  |
Db       5767 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTTCTGG 5826

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
              | |     |||||||||   |||||       |.|||| ||  | |  |     |
Db       5827 AAAGAATCTACGCCCCGGTGCTCGACACGTCGAAAGAGGAACAGCTCAAACTCAGGTACC 5886

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
              |  |     || | ||||  || | |||||||| || || ||||||||||||||   ||
Db       5887 AGATGATGCCCACCGAAGCCAACAAAAGCAGGTACCAGTCTCGAAAAGTAGAAAACCAGA 5946

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
              |||| || ||  | ||||||||  |||| ||  |  |    |||  ||  |  ||
Db       5947 AAGCCATAACCACTGAGCGACTGCTTTCAGGGCTACGGCTGTATAACTCTGCCACAG--- 6003

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
```

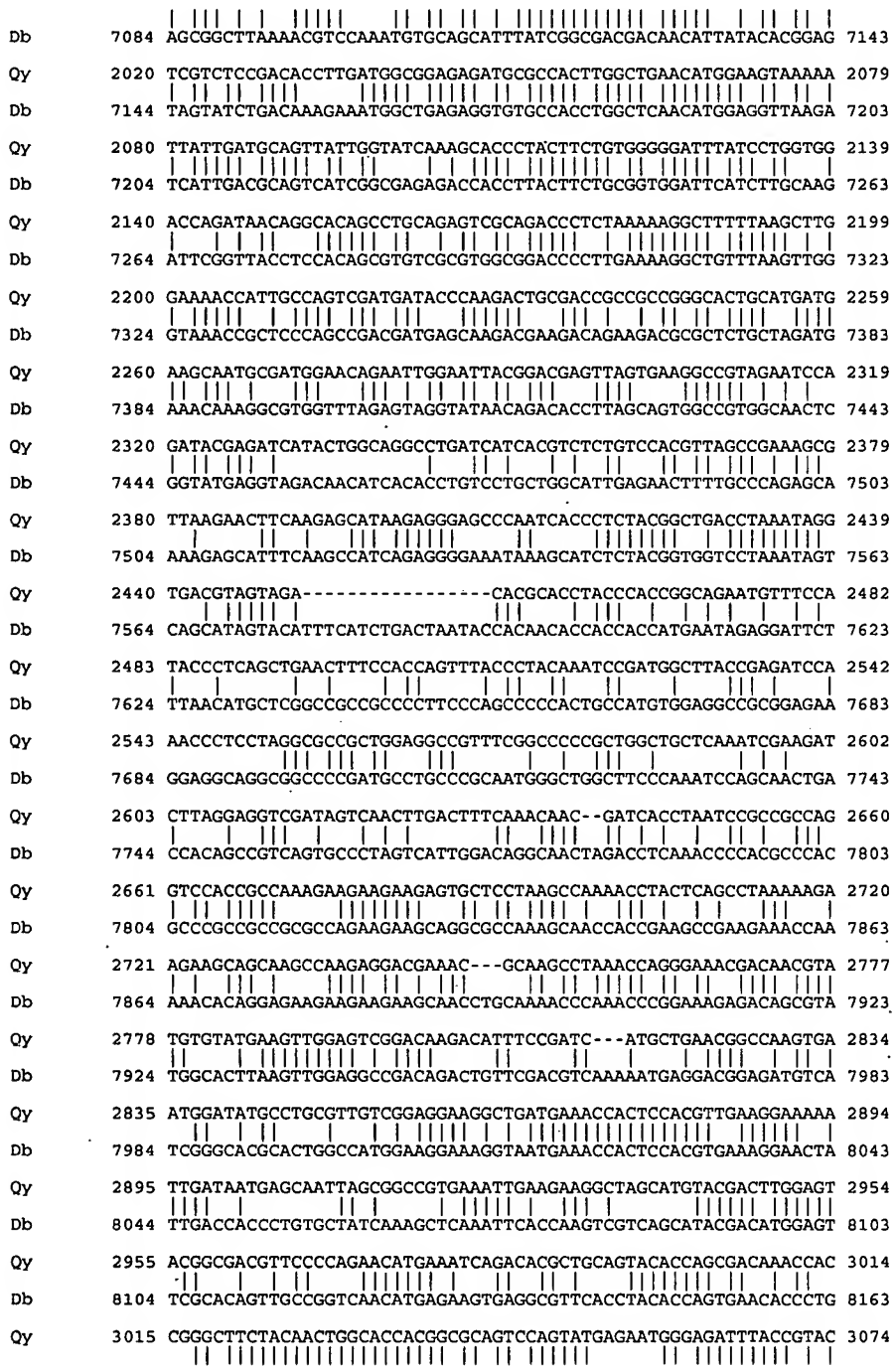
```
                  |||     || || || |   ||| || ||   ||||   || ||    |  |||
Db       6004 ATCAGCCAGAATGCTATAAGATCACCTACCCGAAACCATCGTATTCCAGCAGTGTACCAG 6063

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
              |  |   |      || || | || ||         || |||     | | || |||||||
Db       6064 CGAACTACTCTGACCCAAAGTTTGCTGTAGCTGTTTGTAACAACTATCTGCATGAGAATT 6123

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
              |||| || |||||   ||||   || || || || |||||||| ||  | || ||||| |
Db       6124 ACCCGACGGTAGCATCTTATCAGATCACCGACGAGTACGATGCTTACTTGGATATGGTAG 6183

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
              ||||  ||      || |||||||| || |||||||| || || || || ||||| ||||
Db       6184 ACGGGACAGTCGCTTGCCTAGATACTGCAACTTTTTGCCCCGCCAAGCTTAGAAGTTACC 6243

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
              | || |  ||      |||       ||| | ||  |     || || ||||| | || ||||
Db       6244 CGAAAAGACACGAGTATAGAGCCCCAAACATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6303

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
              | || || ||||| ||  |   ||| || |||||||| || |||||||| || |||||||
Db       6304 ACACGTTGCAAAACGTGCTCATTGCCGCGACTAAAAGAAACTGCAACGTCACACAAATGC 6363

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
              | ||| | ||     | || || ||| |||| || ||||| || ||    ||||| ||||
Db       6364 GTGAACTGCCAACACTGGACTCAGCGACATTCAACGTTGAATGCTTTCGAAAATATGCAT 6423

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
              ||||||| ||||| |||||     |||  |   || || ||| || | ||||| ||| ||
Db       6424 GCAATGACGAGTATTGGGAGGAGTTTGCCCGAAAGCCAATTAGGATCACTACTGAGTTCG 6483

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
              ||||   ||| ||| | |  |||||||| || || || || ||| |||| || || || |
Db       6484 TTACCGCATACGTGGCCAGACTGAAAGGCCCTAAGGCCGCCGCACTGTTCGCAAAGACGC 6543

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
              ||||| |    || ||||| ||  | || ||||||  ||||||||||||  | || ||||
Db       6544 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6603

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
              | || |||||||| || |||||||||||| ||||| ||  | || || || || || | |
Db       6604 ACGTGAAAGTTACACCTGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6663

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
              | || ||||| ||||| || |||||||||||| |||||||||||| ||||| ||||| || |
Db       6664 AAGCCGCAGAACCCCTGGCGACCGCTTACCTATGCGGGATCCACCGGGAGTTAGTGCGCA 6723

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
              | || |  || ||  | || || || || || || ||||| |||||||| ||||| || |
Db       6724 GGCTTACAGCCGTTTTGCTACCCAACATTCACACGCTCTTTGACATGTCGGCGGAGGACT 6783

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
              ||||||| || || || ||||| ||| | || || ||||| ||| |||| |||| |||| |
Db       6784 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGTGACCCGGTACTGGAGACGGATATCG 6843

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
              | ||||| || |||||| |||||||||||||| ||  |  | |   |||||||| | || |
Db       6844 CCTCGTTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGCCTGATGATCTTGGAAG 6903

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
              || | ||||| |||||||| ||  | || ||||| |||   || || ||  | ||  |||
Db       6904 ACCTGGGTGTGGACCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 6963

Qy       1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
              |    || || || || || ||  | || ||||| || || ||||||||||||||| ||||
Db       6964 CCACCCATCTGCCCACGGGTACCCGTTTCAAATTCGGGGCGATGATGAAATCCGGAATGT 7023

Qy       1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
              || | ||||| |||||||||||| |  | ||| || | || || ||||||||| |    |
Db       7024 TCCTCACGCTCTTTGTCAACACAGTTCTGAATGTCGTTATCGCCAGCAGAGTATTGGAGG 7083

Qy       1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
              | ||| | |  |||||    || || || | |||||||||||| |||||  | || || |
```

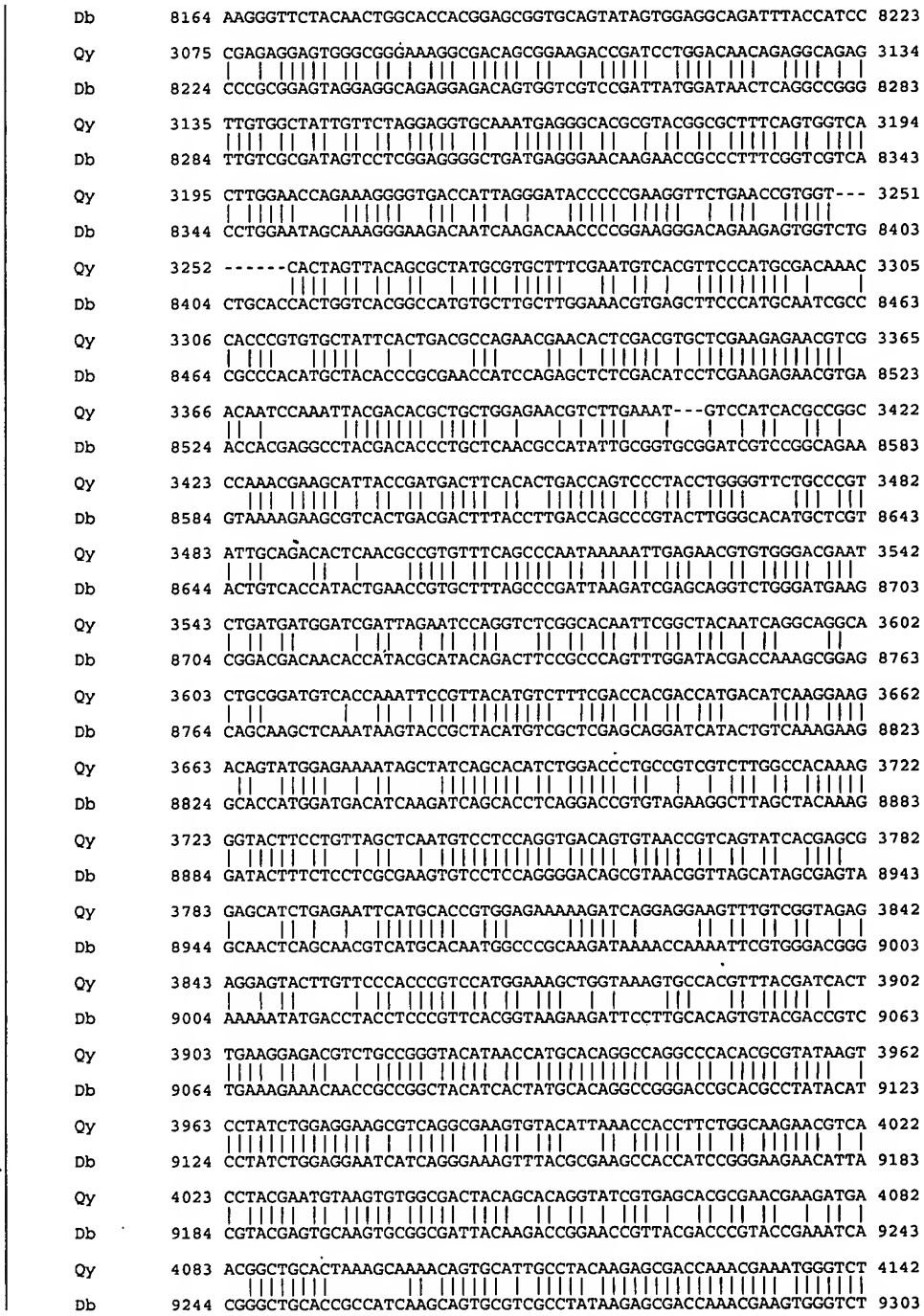


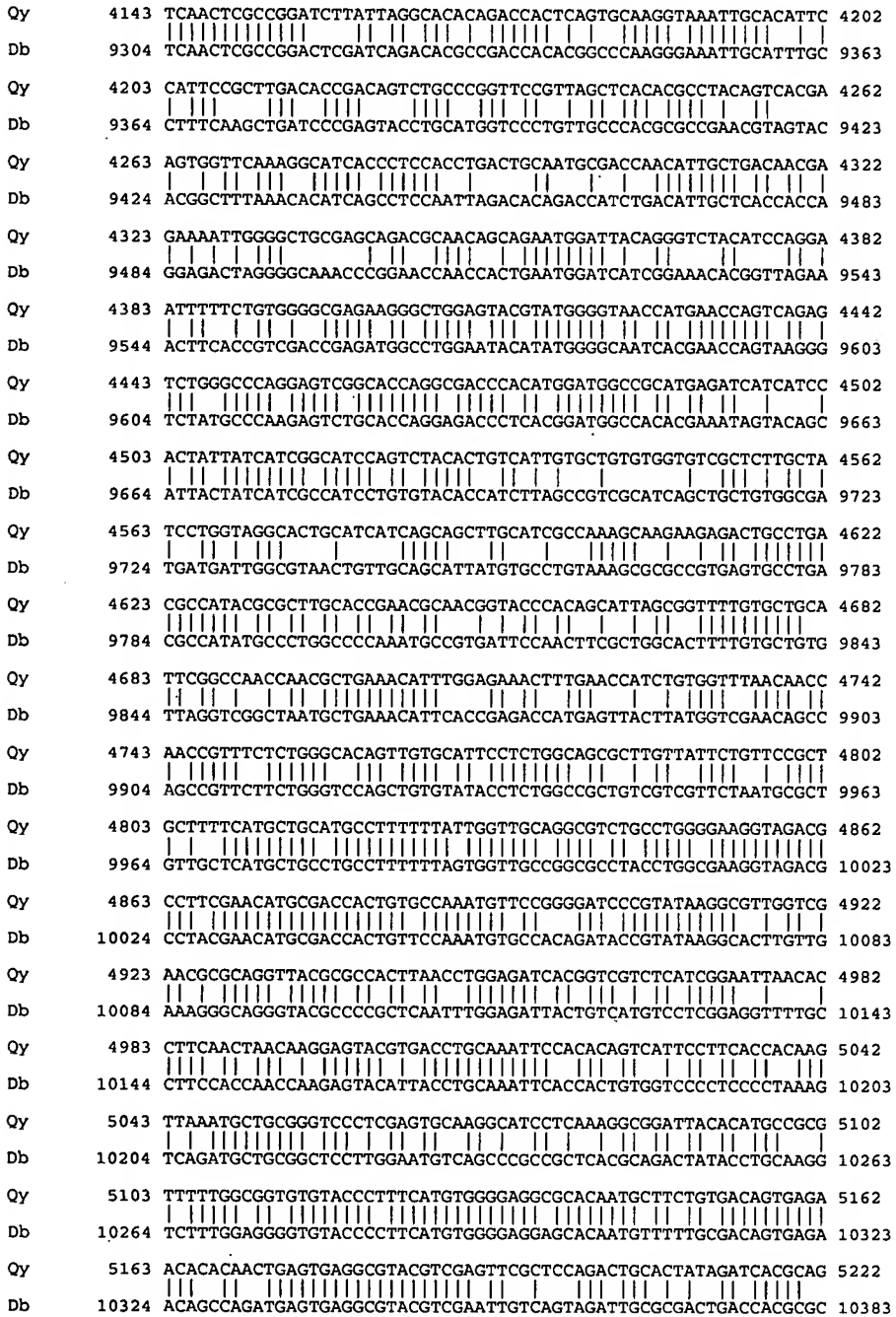

Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9304 TCAACTCGCCGGACTCGATCAGACACGCCGACCACACGGCCCAAGGGAAATTGCATTTGC 9363
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9364 CTTTCAAGCTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAACGTAGTAC 9423
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9424 ACGGCTTTAAACACATCAGCCTCCAATTAGACACAGACCATCTGACATTGCTCACCACCA 9483
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9484 GGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCATCGGAAACACGGTTAGAA 9543
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9544 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGCAATCACGAACCAGTAAGGG 9603
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db 9604 TCTATGCCCAAGAGTCTGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9663
Qy 4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db 9664 ATTACTATCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTGCTGTGGCGA 9723
Qy 4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db 9724 TGATGATTGGCGTAACTGTTGCAGCATTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9783
Qy 4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db 9784 CGCCATATGCCCTGGCCCCAAATGCCGTGATTCCAACTTCGCTGGCACTTTTGTGCTGTG 9843
Qy 4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db 9844 TTAGGTCGGCTAATGCTGAAACATTCACCGAGACCATGAGTTACTTATGGTCGAACAGCC 9903
Qy 4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
Db 9904 AGCCGTTCTTCTGGGTCCAGCTGTGTATACCTCTGGCCGCTGTCGTCGTTCTAATGCGCT 9963
Qy 4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
Db 9964 GTTGCTCATGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10023
Qy 4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
Db 10024 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10083
Qy 4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db 10084 AAAGGGCAGGGTACGCCCCGCTCAATTTGGAGATTACTGTCATGTCCTCGGAGGTTTTGC 10143
Qy 4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db 10144 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCTAAAG 10203
Qy 5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db 10204 TCAGATGCTGCGGCTCCTTGGAATGTCAGCCCGCCGCTCACGCAGACTATACCTGCAAGG 10263
Qy 5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db 10264 TCTTTGGAGGGGTGTACCCCTTCATGTGGGGAGGAGCACAATGTTTTTGCGACAGTGAGA 10323
Qy 5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db 10324 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGTAGATTGCGCGACTGACCACGCGC 10383
Qy 5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db 10384 AGGCGATTAAGGTGCATACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10443
Qy 5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db 10444 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACCTGA 10503
Qy 5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db 10504 AAGTCATAGCTGGACCAATTTCAGCATTGTTTACACCATTCGATCACAAGGTCGTTATCA 10563
Qy 5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db 10564 ATCGCGGCCTGGTGTACAACTATGACTTTCCGGAATACGGAGCGATGAAACCAGGAGCGT 10623
Qy 5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db 10624 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAAGACCTCATCGCCAGCACAGACATTA 10683
Qy 5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db 10684 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCTGGAT 10743
Qy 5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db 10744 TCGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCCCCTTTTGGGTGCA 10803
Qy 5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db 10804 AGATTGCAGTCAATCCGCTTCGAGCGGTGGACTGCTCATACGGGAACATTCCCATTTCTA 10863
Qy 5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db 10864 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10923
Qy 5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db 10924 AATGTGATGTCAGTGAGTGCACTTATTCAGCGGACTTCGGAGGGATGGCTACCCTGCAGT 10983
Qy 5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db 10984 ATGTATCCGACCGCGAAGGACAATGCCCTGTACATTCGCATTCGAGCACAGCAACCCTCC 11043
Qy 5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db 11044 AAGAGTCGACAGTTCATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTCAGCACCGCGA 11103
Qy 5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db 11104 GCCCACAGGCGAACTTCATTGTATCGCTGTGTGGTAAGAAGACAACATGCAATGCAGAAT 11163
Qy 6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db 11164 GCAAACCACCAGCTGATCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTCCAAG 11223
Qy 6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db 11224 CCGCCATCTCAAAAACTTCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11283
Qy 6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
Db 11284 TATTAATTATAGGACTTATGATTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11343
Qy 6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
Db 11344 GACCGCTACGCCCCAATGACCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11403
Qy 6242 GCATAATGCCACGCGCCGCT----------TGACACT 6268
Db 11404 GCATAATGCATCAGGCTGGTATATTAGATCCCCGCTTACCGCGGGCAATATAGCAACACC 11463
Qy 6269 AAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGT 6318

```
Db      11464 AAAACTCGACGTATTTCCGAGGAAGCGCAGTGCATAATGCTGCGCAGTGT 11513


RESULT 4
US-09-367-764-7
; Sequence 7, Application US/09367764
; Patent No. 6583121
;  GENERAL INFORMATION:
;    APPLICANT:  Johnston, Robert E.
;    APPLICANT:  Davis, Nancy L.
;    APPLICANT:  Simpson, Dennis A.
;    TITLE OF INVENTION:  System for the In Vivo Delivery and
;    TITLE OF INVENTION:  Expression of Heterologous Genes in the Bone Marrow
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson, P.A.
;      STREET:  1211 East Morehead Street
;      CITY:  Charlotte
;      STATE:  No. 6583121th Carolina
;      COUNTRY:  USA
;      ZIP:  28234
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/367,764
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/801,263
;      FILING DATE:  19-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sibley, Kenneth D.
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5470-147
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-420-2200
;      TELEFAX:  919-881-3175
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11663 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-367-764-7

  Query Match             37.0%;  Score 2397.6;  DB 4;  Length 11663;
  Best Local Similarity   64.6%;  Pred. No. 0;
  Matches 3752;  Conservative    0;  Mismatches 1984;  Indels   74;  Gaps    9;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
              |||| |     |  |  || | ||| || || ||  |||| |||||||  || ||| | |
Db       5707 ACTGTCTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5766

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
              || |||| || || |  ||      | ||   |||||| |  ||||| ||     || | |
Db       5767 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTTCTGG 5826

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
              | |     |||||||||   ||||       | |||| ||  | |  |     |     |
Db       5827 AAAGAATCTACGCCCCGGTGCTCGACACGTCGAAAGAGGAACAGCTCAAACTCAGGTACC 5886

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
              | |     || | ||||  || | |||||||||| || || ||||||||||||||   ||
Db       5887 AGATGATGCCCACCGAAGCCAACAAAAGCAGGTACCAGTCTCGAAAAGTAGAAAACCAGA 5946

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
              |||| || ||  | |||||||| |||| ||  |  |     |||  ||  |   ||
Db       5947 AAGCCATAACCACTGAGCGACTGCTTTCAGGGCTACGGCTGTATAACTCTGCCACAG--- 6003
```

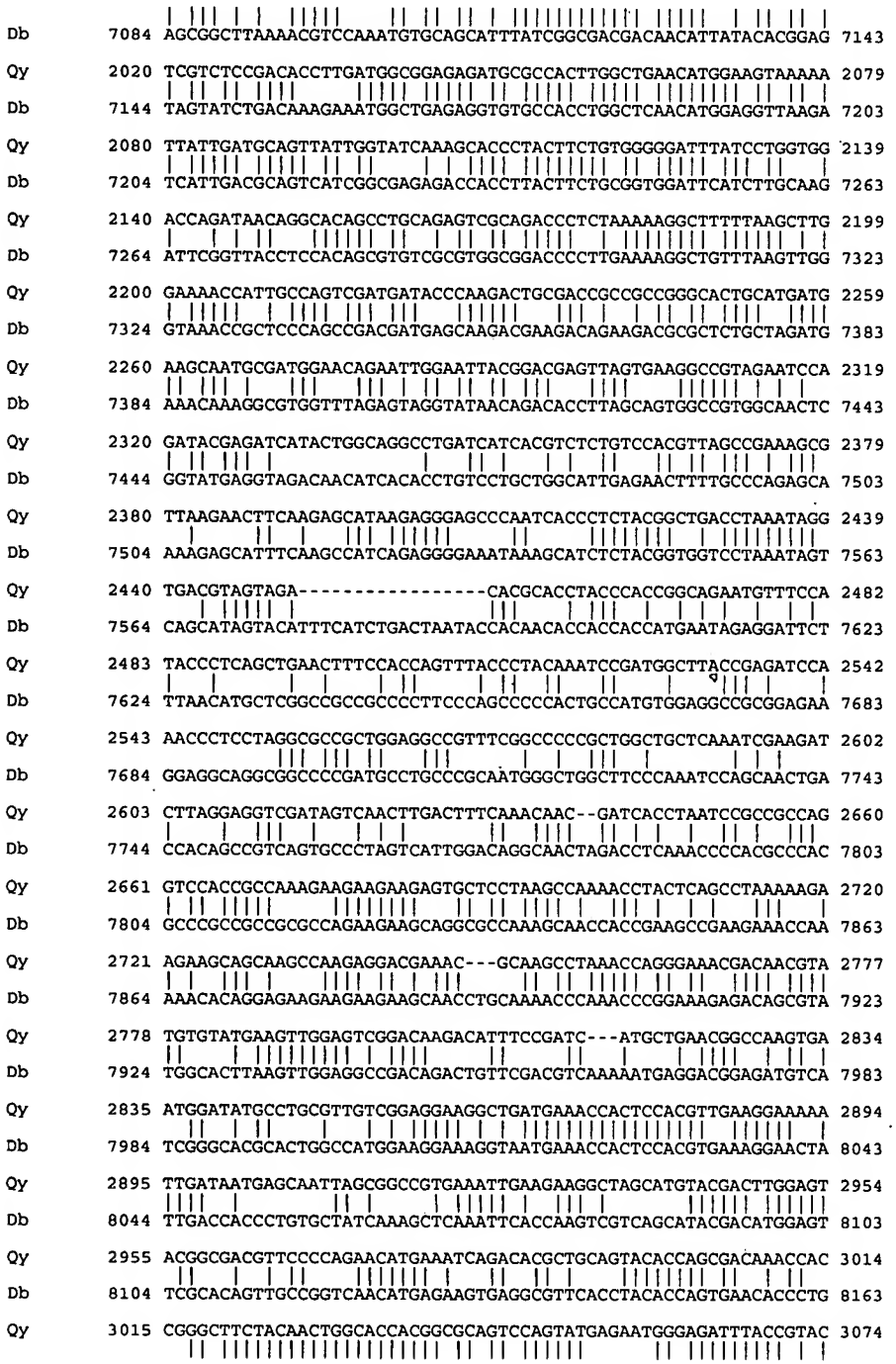
```
Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
              |||     || || || |   ||| || ||  ||||  || ||     |  |||
Db       6004 ATCAGCCAGAATGCTATAAGATCACCTACCCGAAACCATCGTATTCCAGCAGTGTACCAG 6063

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
              |  |   |     || || | || ||       || |||    | | || |||||||
Db       6064 CGAACTACTCTGACCCAAAGTTTGCTGTAGCTGTTTGTAACAACTATCTGCATGAGAATT 6123

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
              |||| || |||||   ||||  || || || || |||||||| ||  | || ||||| |
Db       6124 ACCCGACGGTAGCATCTTATCAGATCACCGACGAGTACGATGCTTACTTGGATATGGTAG 6183

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
              ||||  ||      || |||||||| || |||||||| || || || || ||||| ||||
Db       6184 ACGGGACAGTCGCTTGCCTAGATACTGCAACTTTTTGCCCCGCCAAGCTTAGAAGTTACC 6243

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
              | || |  ||    |||       ||| | ||  |    || || |||||  | || ||||
Db       6244 CGAAAAGACACGAGTATAGAGCCCCAAACATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6303

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
              | || || ||||| ||  |   ||| || |||||||| || |||||||| || |||||||
Db       6304 ACACGTTGCAAAACGTGCTCATTGCCGCGACTAAAAGAAACTGCAACGTCACACAAATGC 6363

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
              | ||| | ||      | || || |||| || || ||||| || ||    ||||| ||||
Db       6364 GTGAACTGCCAACACTGGACTCAGCGACATTCAACGTTGAATGCTTTCGAAAATATGCAT 6423

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
              ||||||| ||||| |||||    ||| |  || || ||| || | |||||  |||  ||
Db       6424 GCAATGACGAGTATTGGGAGGAGTTTGCCCGAAAGCCAATTAGGATCACTACTGAGTTCG 6483

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
              ||||   ||| ||| | |  |||||||| || || || || ||| |||| || || || |
Db       6484 TTACCGCATACGTGGCCAGACTGAAAGGCCCTAAGGCCGCCGCACTGTTCGCAAAGACGC 6543

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
              |||||  |    || ||||| ||  | || ||||||  ||||||||||||  | || ||||
Db       6544 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6603

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
              | || |||||||| || |||||||||||| ||||| ||  | || || || || || || |
Db       6604 ACGTGAAAGTTACACCTGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6663

Qy       1540 AGGCTGCAGATCCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
              | || ||||| ||||| || |||||||||||| |||||||||||| ||||| ||||| || |
Db       6664 AAGCCGCAGAACCCCTGGCGACCGCTTACCTATGCGGGATCCACCGGGAGTTAGTGCGCA 6723

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
              | || |  || ||  | || || || || || ||||| |||||||| ||||| || |
Db       6724 GGCTTACAGCCGTTTTGCTACCCAACATTCACACGCTCTTTGACATGTCGGCGGAGGACT 6783

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
              ||||||| || || || ||||| ||| | || || ||||| ||| |||| ||||| ||||
Db       6784 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGTGACCCGGTACTGGAGACGGATATCG 6843

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
              | ||||| || |||||| |||||||||||||| ||  |  | |  |||||||| | || |
Db       6844 CCTCGTTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGCCTGATGATCTTGGAAG 6903

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
              || | ||||| |||||||| || | || |||| |||  ||  || || ||  | ||  |||
Db       6904 ACCTGGGTGTGGACCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 6963

Qy       1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
              |    || || || || || ||  | || ||||| || || ||||||||||||||  ||||
Db       6964 CCACCCATCTGCCCACGGGTACCCGTTTCAAATTCGGGGCGATGATGAAATCCGGAATGT 7023

Qy       1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
              || | ||||| |||||||||||| |  | ||| || | || || |||||||| |    |
Db       7024 TCCTCACGCTCTTTGTCAACACAGTTCTGAATGTCGTTATCGCCAGCAGAGTATTGGAGG 7083

Qy       1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
```

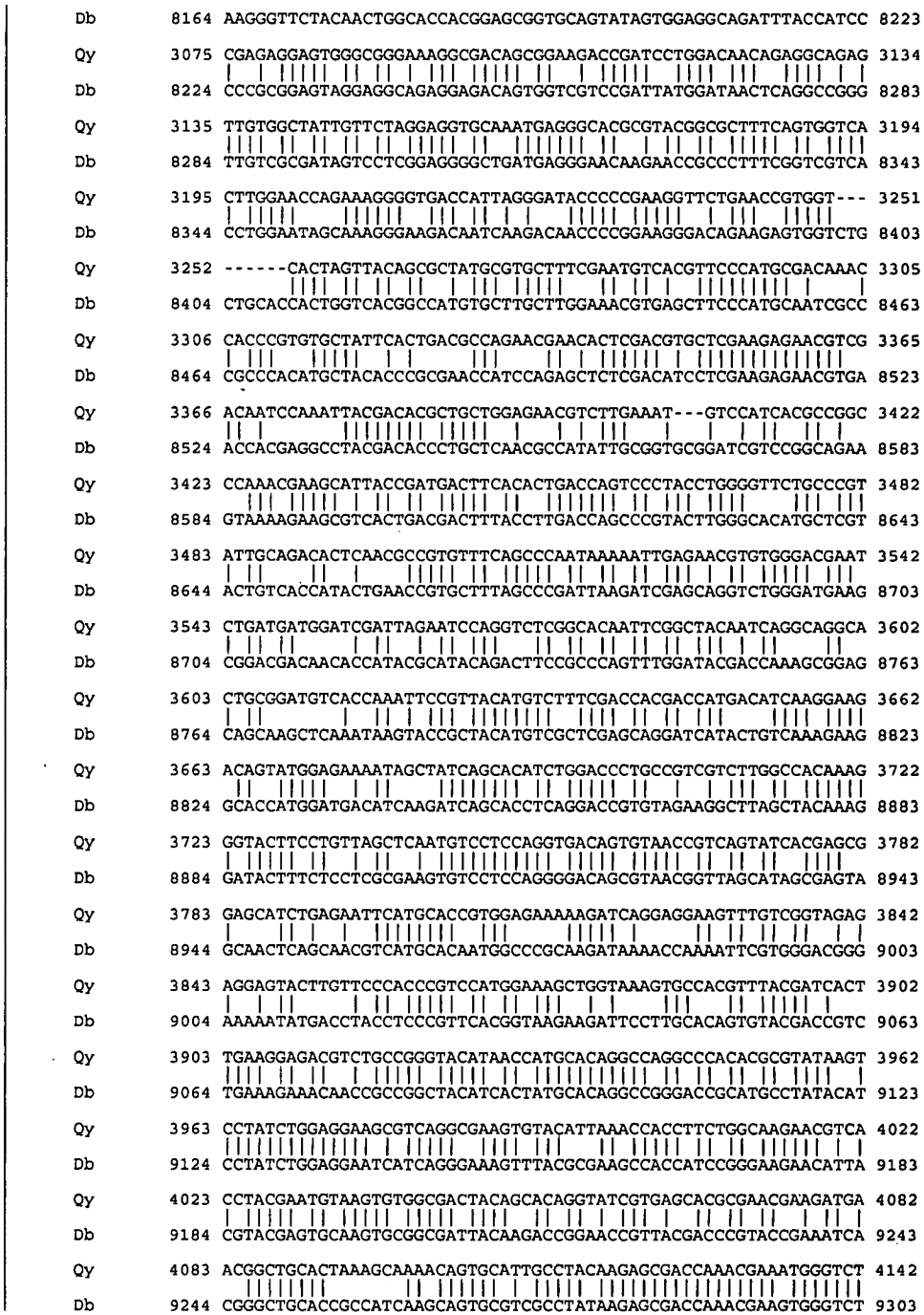


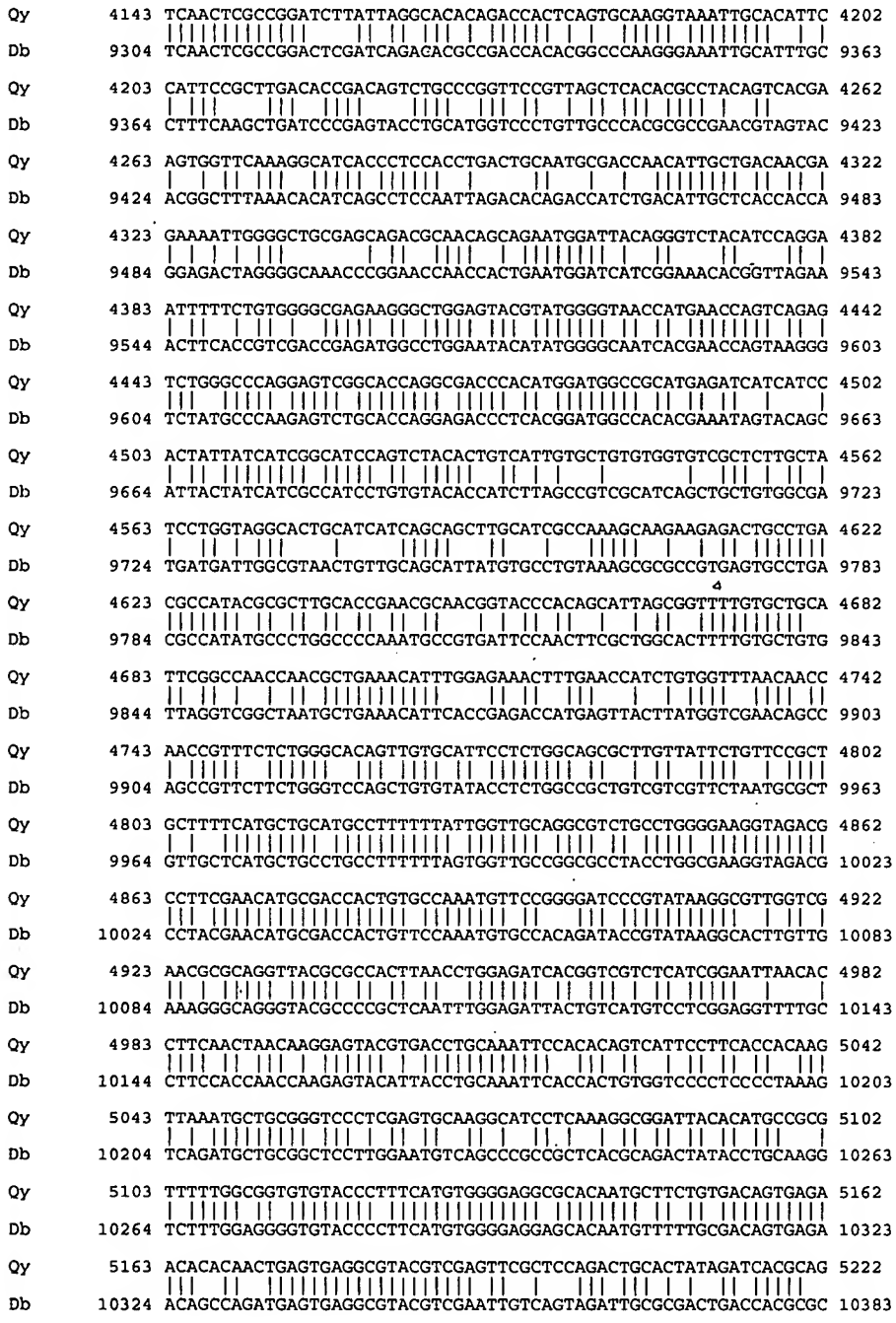



Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9304 TCAACTCGCCGGACTCGATCAGACACGCCGACCACACGGCCCAAGGGAAATTGCATTTGC 9363
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9364 CTTTCAAGCTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAACGTAGTAC 9423
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9424 ACGGCTTTAAACACATCAGCCTCCAATTAGACACAGACCATCTGACATTGCTCACCACCA 9483
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9484 GGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCATCGGAAACACGGTTAGAA 9543
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9544 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGCAATCACGAACCAGTAAGGG 9603
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db 9604 TCTATGCCCAAGAGTCTGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9663
Qy 4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db 9664 ATTACTATCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTGCTGTGGCGA 9723
Qy 4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db 9724 TGATGATTGGCGTAACTGTTGCAGCATTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9783
Qy 4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db 9784 CGCCATATGCCCTGGCCCCAAATGCCGTGATTCCAACTTCGCTGGCACTTTTGTGCTGTG 9843
Qy 4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db 9844 TTAGGTCGGCTAATGCTGAAACATTCACCGAGACCATGAGTTACTTATGGTCGAACAGCC 9903
Qy 4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
Db 9904 AGCCGTTCTTCTGGGTCCAGCTGTGTATACCTCTGGCCGCTGTCGTCGTTCTAATGCGCT 9963
Qy 4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
Db 9964 GTTGCTCATGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10023
Qy 4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
Db 10024 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10083
Qy 4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db 10084 AAAGGGCAGGGTACGCCCCGCTCAATTTGGAGATTACTGTCATGTCCTCGGAGGTTTTGC 10143
Qy 4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db 10144 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCTAAAG 10203
Qy 5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db 10204 TCAGATGCTGCGGCTCCTTGGAATGTCAGCCCGCCGCTCACGCAGACTATACCTGCAAGG 10263
Qy 5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db 10264 TCTTTGGAGGGGTGTACCCCTTCATGTGGGGAGGAGCACAATGTTTTTGCGACAGTGAGA 10323
Qy 5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db 10324 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGTAGATTGCGCGACTGACCACGCGC 10383
Qy 5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db 10384 AGGCGATTAAGGTGCATACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10443
Qy 5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db 10444 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACCTGA 10503
Qy 5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db 10504 AAGTCATAGCTGGACCAATTTCAGCATTGTTTACACCATTCGATCACAAGGTCGTTATCA 10563
Qy 5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db 10564 ATCGCGGCCTGGTGTACAACTATGACTTTCCGGAATACGGAGCGATGAAACCAGGAGCGT 10623
Qy 5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db 10624 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAAGACCTCATCGCCAGCACAGACATTA 10683
Qy 5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db 10684 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCTGGAT 10743
Qy 5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db 10744 TCGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCCCCTTTTGGGTGCA 10803
Qy 5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db 10804 AGATTGCAGTCAATCCGCTTCGAGCGGTGGACTGCTCATACGGGAACATTCCCATTTCTA 10863
Qy 5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db 10864 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10923
Qy 5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db 10924 AATGTGATGTCAGTGAGTGCACTTATTCAGCGGACTTCGGAGGGATGGCTACCCTGCAGT 10983
Qy 5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db 10984 ATGTATCCGACCGCGAAGGACAATGCCCTGTACATTCGCATTCGAGCACAGCAACCCTCC 11043
Qy 5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db 11044 AAGAGTCGACAGTTCATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTCAGCACCGCGA 11103
Qy 5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db 11104 GCCCACAGGCGAACTTCATTGTATCGCTGTGTGGTAAGAAGACAACATGCAATGCAGAAT 11163
Qy 6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db 11164 GCAAACCACCAGCTGATCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTCCAAG 11223
Qy 6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db 11224 CCGCCATCTCAAAAACTTCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11283
Qy 6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
Db 11284 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11343
Qy 6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
Db 11344 GACCGCTACGCCCCAATGACCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11403
Qy 6242 GCATAATGCCACGCGCCGCT----------TGACACT 6268
Db 11404 GCATAATGCATCAGGCTGGTATATTAGATCCCCGCTTACCGCGGGCAATATAGCAACACC 11463
Qy 6269 AAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGT 6318

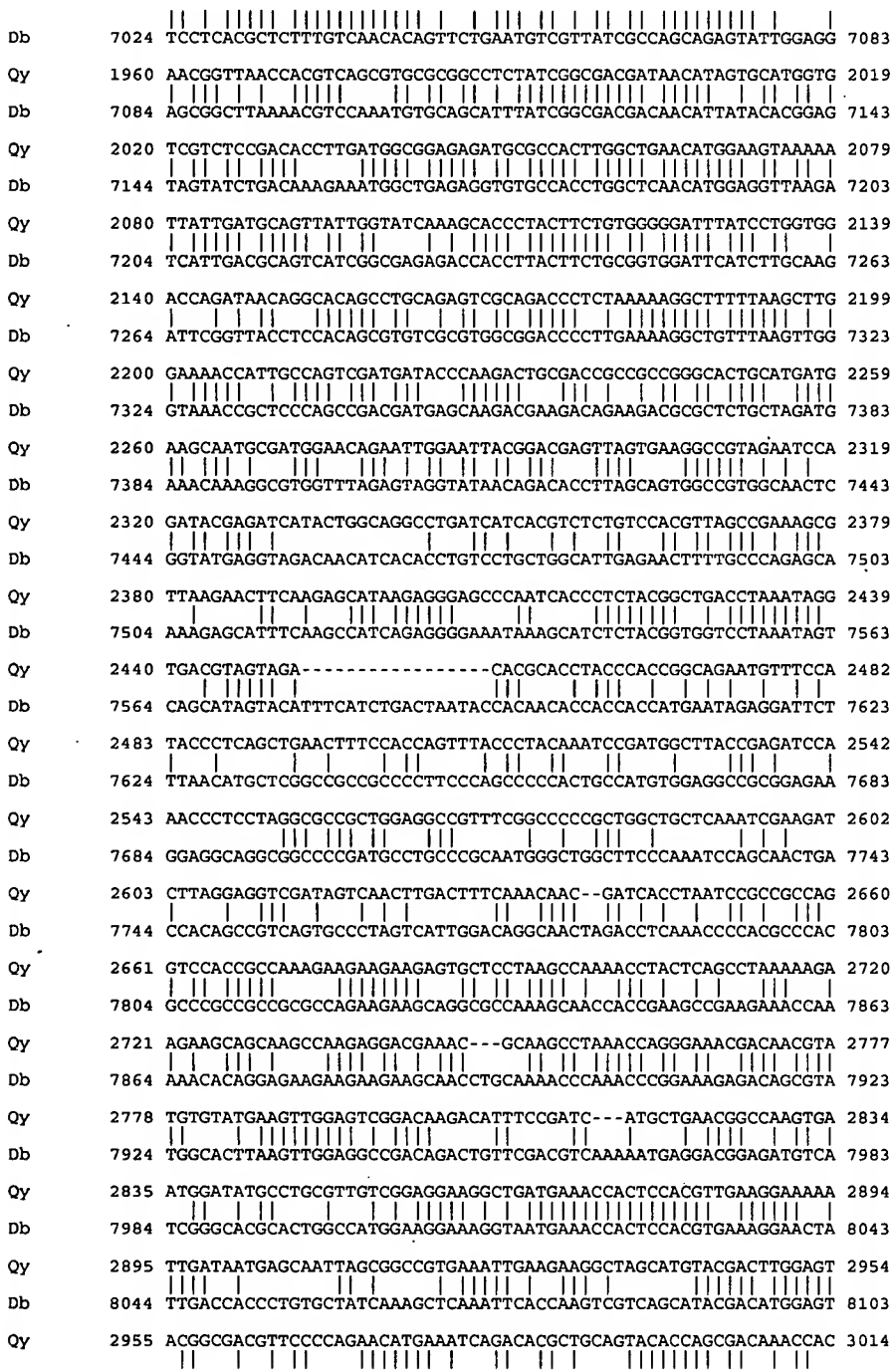
```
                 ||||||||| |||||||||||||||| |||||||||||||| |||||||
Db      11464 AAAACTCGACGTATTTCCGAGGAAGCGCAGTGCATAATGCTGCGCAGTGT 11513


RESULT 5
US-08-446-932-1
; Sequence 1, Application US/08446932
; Patent No. 5639650
;  GENERAL INFORMATION:
;    APPLICANT:  Johnston, Robert E.
;    APPLICANT:  Simpson, Dennis
;    APPLICANT:  Davis, Nancy L.
;    TITLE OF INVENTION:  cDNA Clone for South African
;    TITLE OF INVENTION:  Arbovirus No. 5639650 86
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kenneth D. Sibley
;      STREET:  Post Office Drawer 34009
;      CITY:  Charlotte
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  28234
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/446,932
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sibley, Kenneth D.
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5470-118
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 881-3140
;      TELEFAX:  (919) 881-3175
;      TELEX:  575102
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11663 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
US-08-446-932-1

  Query Match             36.9%;  Score 2396;  DB 1;  Length 11663;
  Best Local Similarity   64.6%;  Pred. No. 0;
  Matches 3751;  Conservative    0;  Mismatches 1985;  Indels   74;  Gaps    9;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
              |||| |     |  |  || | ||| || || ||  |||| |||||||  || ||| | |
Db       5707 ACTGTCTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5766

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
              || |||| || || |  ||      | ||   |||||| |  ||||| ||    || | |
Db       5767 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTTCTGG 5826

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
              | |     |||||||||   |||||        | |||| ||  | |  |      |
Db       5827 AAAGAATCTACGCCCCGGTGCTCGACACGTCGAAAGAGGAACAGCTCAAACTCAGGTACC 5886

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
              |  |      || | ||||  || | |||||||||| || || ||||||||||||||   ||
Db       5887 AGATGATGCCCACCGAAGCCAACAAAAGCAGGTACCAGTCTCGAAAAGTAGAAAACCAGA 5946

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
              |||| || ||  | |||||||| |||| ||  |  |    |||  ||  |   ||
Db       5947 AAGCCATAACCACTGAGCGACTGCTTTCAGGGCTACGACTGTATAACTCTGCCACAG--- 6003

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
              |||      || || || |   ||| || ||  ||||  || ||     | |||
Db       6004 ATCAGCCAGAATGCTATAAGATCACCTACCCGAAACCATCGTATTCCAGCAGTGTACCAG 6063

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
              |  |  |     || || | || ||  .  || |||     | | || |||||||
Db       6064 CGAACTACTCTGACCCAAAGTTTGCTGTAGCTGTTTGTAACAACTATCTGCATGAGAATT 6123

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
              |||| || |||||    ||||   || || || || |||||||| ||  | || ||||| |
Db       6124 ACCCGACGGTAGCATCTTATCAGATCACCGACGAGTACGATGCTTACTTGGATATGGTAG 6183

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
              ||||  ||       || |||||||| || |||||||| || || || || ||||| ||||
Db       6184 ACGGGACAGTCGCTTGCCTAGATACTGCAACTTTTTGCCCCGCCAAGCTTAGAAGTTACC 6243

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
              | || |  ||     |||      ||| | ||  |    || || |||||  | || ||||
Db       6244 CGAAAAGACACGAGTATAGAGCCCCAAACATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6303

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
              | || || ||||| ||  |   ||| || |||||||| || |||||||| || |||||||
Db       6304 ACACGTTGCAAAACGTGCTCATTGCCGCGACTAAAAGAAACTGCAACGTCACACAAATGC 6363

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
              | ||| | || \   | || || ||| |||| || ||||| || ||    ||||| ||||
Db       6364 GTGAACTGCCAACACTGGACTCAGCGACATTCAACGTTGAATGCTTTCGAAAATATGCAT 6423

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
              ||||||| ||||| |||||     |||  |   || || ||| || | ||||| |||  ||
Db       6424 GCAATGACGAGTATTGGGAGGAGTTTGCCCGAAAGCCAATTAGGATCACTACTGAGTTCG 6483

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
              ||||   ||| ||| | |  |||||||| || || || || ||| |||| || || || |
Db       6484 TTACCGCATACGTGGCCAGACTGAAAGGCCCTAAGGCCGCCGCACTGTTCGCAAAGACGC 6543

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
              ||||| |    || ||||| ||  | || ||||||  |||||||||||  | || ||||
Db       6544 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6603

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
              | || |||||||| || ||||||||||| ||||| ||  | || || || || || || |
Db       6604 ACGTGAAAGTTACACCTGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6663

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
              | || ||||| ||||| || |||||||||| |||||||||| ||||| ||||| || |
Db       6664 AAGCCGCAGAACCCCTGGCGACCGCTTACCTATGCGGGATCCACCGGGAGTTAGTGCGCA 6723

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
              | || |  || || || || || || || ||||  |||||||| ||||| || |
Db       6724 GGCTTACAGCCGTTTTGCTACCCAACATTCACACGCTCTTTGACATGTCGGCGGAGGACT 6783

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
              ||||||| || || || ||||| ||| | || || ||||| ||| |||| ||||| ||||
Db       6784 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGTGACCCGGTACTGGAGACGGATATCG 6843

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
              | ||||| || |||||| |||||||||||| ||  |  | |  |||||||  | || |
Db       6844 CCTCGTTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGCCTGATGATCTTGGAAG 6903

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
              || | ||||| |||||||| ||  | || ||||| |||   || || || |  || |||
Db       6904 ACCTGGGTGTGGACCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 6963

Qy       1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
              |     || || || || || || |  | || ||||| || || ||||||||||||| ||||
Db       6964 CCACCCATCTGCCCACGGGTACCCGTTTCAAATTCGGGGCGATGATGAAATCCGGAATGT 7023

Qy       1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
              || | ||||| |||||||||||| |  | ||| || | || || |||||||| |   |
Db       7024 TCCTCACGCTCTTTGTCAACACAGTTCTGAATGTCGTTATCGCCAGCAGAGTATTGGAGG 7083

Qy       1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
```

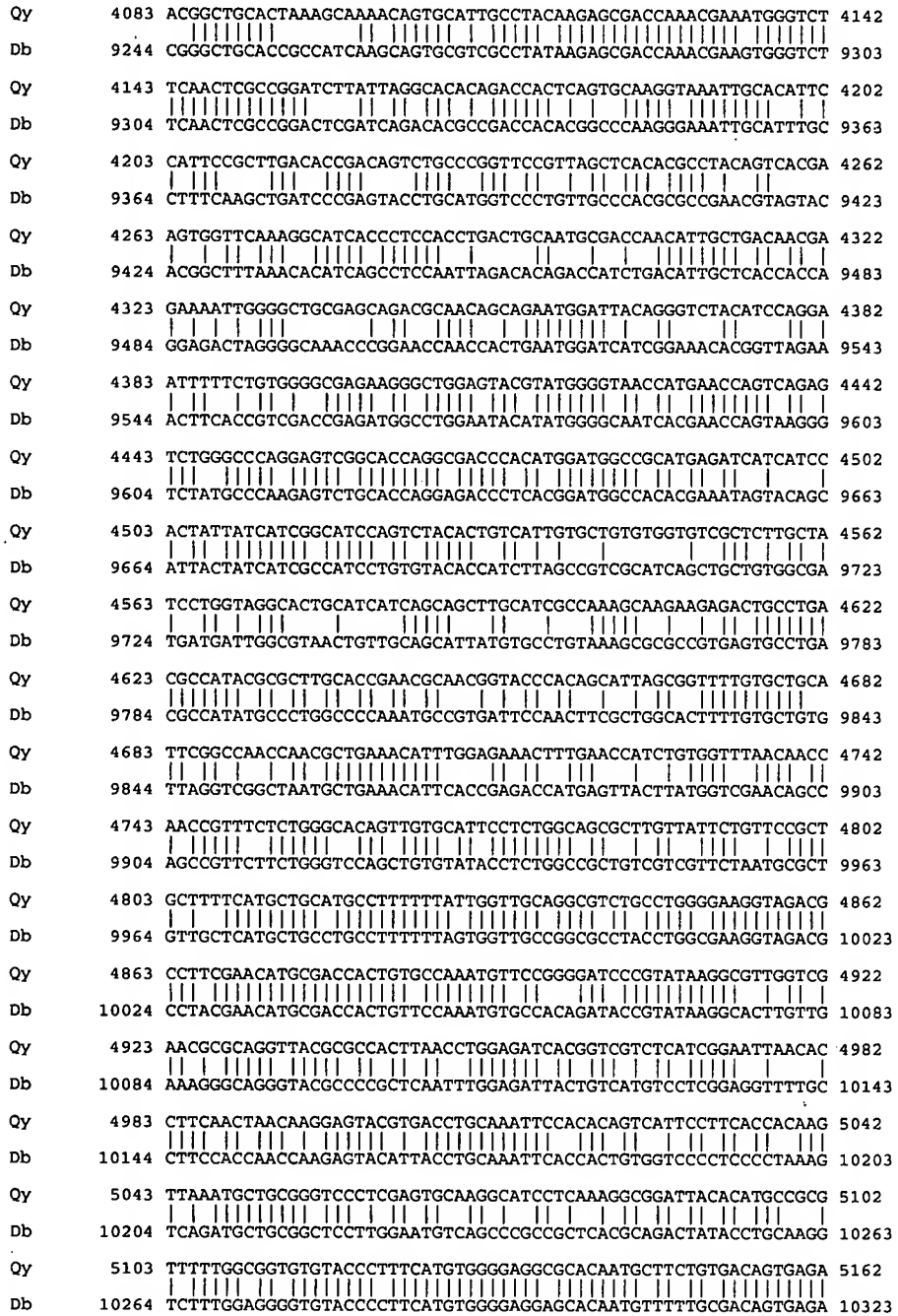


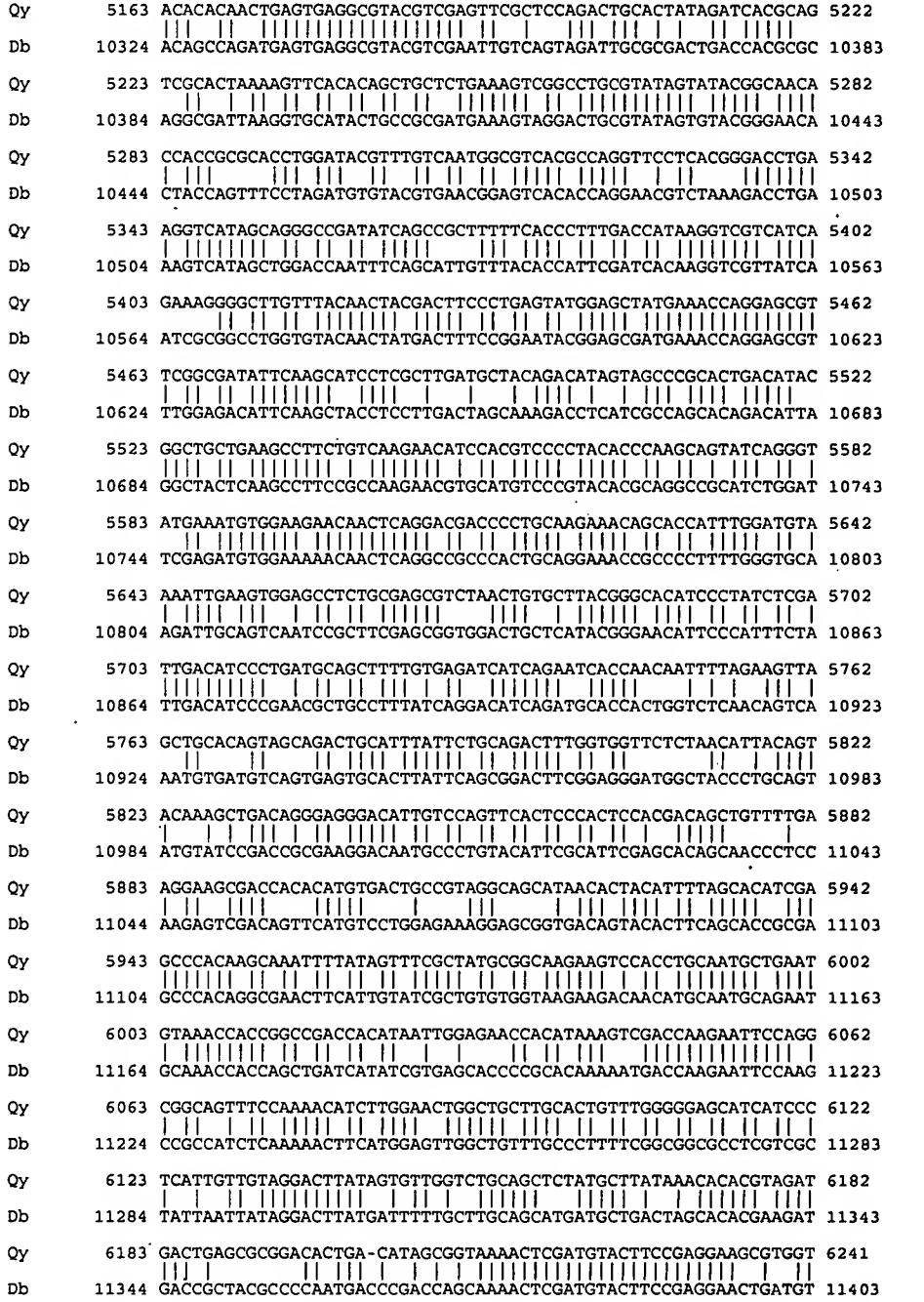



Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9304 TCAACTCGCCGGACTCGATCAGACACGCCGACCACACACGGCCCAAGGGAAATTGCATTTGC 9363
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9364 CTTTCAAGCTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAACGTAGTAC 9423
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9424 ACGGCTTTAAACACATCAGCCTCCAATTAGACACAGACCATCTGACATTGCTCACCACCA 9483
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9484 GGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCATCGGAAACACGGTTAGAA 9543
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9544 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGCAATCACGAACCAGTAAGGG 9603
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db 9604 TCTATGCCCAAGAGTCTGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9663
Qy 4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db 9664 ATTACTATCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTGCTGTGGCGA 9723
Qy 4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db 9724 TGATGATTGGCGTAACTGTTGCAGCATTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9783
Qy 4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db 9784 CGCCATATGCCCTGGCCCCAAATGCCGTGATTCCAACTTCGCTGGCACTTTTGTGCTGTG 9843
Qy 4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db 9844 TTAGGTCGGCTAATGCTGAAACATTCACCGAGACCATGAGTTACTTATGGTCGAACAGCC 9903
Qy 4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
Db 9904 AGCCGTTCTTCTGGGTCCAGCTGTGTATACCTCTGGCCGCTGTCGTCGTTCTAATGCGCT 9963
Qy 4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
Db 9964 GTTGCTCATGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10023
Qy 4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
Db 10024 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10083
Qy 4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db 10084 AAAGGGCAGGGTACGCCCCGCTCAATTTGGAGATTACTGTCATGTCCTCGGAGGTTTTGC 10143
Qy 4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db 10144 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCTAAAG 10203
Qy 5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db 10204 TCAGATGCTGCGGCTCCTTGGAATGTCAGCCCGCCGCTCACGCAGACTATACCTGCAAGG 10263
Qy 5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db 10264 TCTTTGGAGGGGTGTACCCCTTCATGTGGGGAGGAGCACAATGTTTTTGCGACAGTGAGA 10323
Qy 5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db 10324 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGTAGATTGCGCGACTGACCACGCGC 10383
Qy 5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db 10384 AGGCGATTAAGGTGCATACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10443
Qy 5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db 10444 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACCTGA 10503
Qy 5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db 10504 AAGTCATAGCTGGACCAATTTCAGCATTGTTTACACCATTCGATCACAAGGTCGTTATCA 10563
Qy 5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db 10564 ATCGCGGCCTGGTGTACAACTATGACTTTCCGGAATACGGAGCGATGAAACCAGGAGCGT 10623
Qy 5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db 10624 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAAGACCTCATCGCCAGCACAGACATTA 10683
Qy 5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db 10684 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCTGGAT 10743
Qy 5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db 10744 TCGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCCCCTTTTGGGTGCA 10803
Qy 5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db 10804 AGATTGCAGTCAATCCGCTTCGAGCGGTGGACTGCTCATACGGGAACATTCCCATTTCTA 10863
Qy 5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db 10864 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10923
Qy 5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db 10924 AATGTGATGTCAGTGAGTGCACTTATTCAGCGGACTTCGGAGGGATGGCTACCCTGCAGT 10983
Qy 5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db 10984 ATGTATCCGACCGCGAAGGACAATGCCCTGTACATTCGCATTCGAGCACAGCAACCCTCC 11043
Qy 5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db 11044 AAGAGTCGACAGTTCATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTCAGCACCGCGA 11103
Qy 5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db 11104 GCCCACAGGCGAACTTCATTGTATCGCTGTGTGGTAAGAAGACAACATGCAATGCAGAAT 11163
Qy 6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db 11164 GCAAACCACCAGCTGATCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTCCAAG 11223
Qy 6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db 11224 CCGCCATCTCAAAAACTTCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11283
Qy 6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
Db 11284 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11343
Qy 6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
Db 11344 GACCGCTACGCCCCAATGACCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11403
Qy 6242 GCATAATGCCACGCGCCGCT----------TGACACT 6268
Db 11404 GCATAATGCATCAGGCTGGTATATTAGATCCCCGCTTACCGCGGGCAATATAGCAACACC 11463
Qy 6269 AAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGT 6318

```
Db      11464 AAAACTCGACGTATTTCCGAGGAAGCGCAGTGCATAATGCTGCGCAGTGT 11513

RESULT 6
US-08-801-263A-1
; Sequence 1, Application US/08801263A
; Patent No. 5811407
;  GENERAL INFORMATION:
;    APPLICANT:  Johnston, Robert E.
;    APPLICANT:  Davis, Nancy L.
;    APPLICANT:  Simpson, Dennis A.
;    TITLE OF INVENTION:  System for the In Vivo Delivery and
;    TITLE OF INVENTION:  Expression of Heterologous Genes in the Bone Marrow
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson, P.A.
;      STREET:  1211 East Morehead Street
;      CITY:  Charlotte
;      STATE:  No. 5811407th Carolina
;      COUNTRY:  USA
;      ZIP:  28234
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/801,263A
;      FILING DATE:  19-FEB-1997
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sibley, Kenneth D.
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5470-147
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-420-2200
;      TELEFAX:  919-881-3175
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11663 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  60..7559
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  7608..11342
US-08-801-263A-1

  Query Match             36.9%;  Score 2396;  DB 1;  Length 11663;
  Best Local Similarity   64.6%;  Pred. No. 0;
  Matches 3751;  Conservative    0;  Mismatches 1985;  Indels   74;  Gaps    9;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
Db       5707 ACTGTCTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5766

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
Db       5767 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTTCTGG 5826

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
Db       5827 AAAGAATCTACGCCCCGGTGCTCGACACGTCGAAAGAGGAACAGCTCAAACTCAGGTACC 5886

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
Db       5887 AGATGATGCCCACCGAAGCCAACAAAAGCAGGTACCAGTCTCGAAAAGTAGAAAACCAGA 5946

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
Db       5947 AAGCCATAACCACTGAGCGACTGCTTTCAGGGCTACGACTGTATAACTCTGCCACAG--- 6003

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
Db       6004 ATCAGCCAGAATGCTATAAGATCACCTACCCGAAACCATCGTATTCCAGCAGTGTACCAG 6063

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
Db       6064 CGAACTACTCTGACCCAAAGTTTGCTGTAGCTGTTTGTAACAACTATCTGCATGAGAATT 6123

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
Db       6124 ACCCGACGGTAGCATCTTATCAGATCACCGACGAGTACGATGCTTACTTGGATATGGTAG 6183

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
Db       6184 ACGGGACAGTCGCTTGCCTAGATACTGCAACTTTTTGCCCCGCCAAGCTTAGAAGTTACC 6243

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
Db       6244 CGAAAAGACACGAGTATAGAGCCCCAAACATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6303

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
Db       6304 ACACGTTGCAAAACGTGCTCATTGCCGCGACTAAAAGAAACTGCAACGTCACACAAATGC 6363

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
Db       6364 GTGAACTGCCAACACTGGACTCAGCGACATTCAACGTTGAATGCTTTCGAAAATATGCAT 6423

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
Db       6424 GCAATGACGAGTATTGGGAGGAGTTTGCCCGAAAGCCAATTAGGATCACTACTGAGTTCG 6483

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
Db       6484 TTACCGCATACGTGGCCAGACTGAAAGGCCCTAAGGCCGCCGCACTGTTCGCAAAGACGC 6543

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
Db       6544 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6603

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
Db       6604 ACGTGAAAGTTACACCTGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6663

Qy       1540 AGGCTGCAGATCCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
Db       6664 AAGCCGCAGAACCCCTGGCGACCGCTTACCTATGCGGGATCCACCGGGAGTTAGTGCGCA 6723

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
Db       6724 GGCTTACAGCCGTTTTGCTACCCAACATTCACACGCTCTTTGACATGTCGGCGGAGGACT 6783

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
Db       6784 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGTGACCCGGTACTGGAGACGGATATCG 6843

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
Db       6844 CCTCGTTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGCCTGATGATCTTGGAAG 6903

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
Db       6904 ACCTGGGTGTGGACCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 6963

Qy       1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
Db       6964 CCACCCATCTGCCCACGGGTACCCGTTTCAAATTCGGGGCGATGATGAAATCCGGAATGT 7023

Qy       1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
```


Db 7024 TCCTCACGCTCTTTGTCAACACAGTTCTGAATGTCGTTATCGCCAGCAGAGTATTGGAGG 7083
Qy 1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
Db 7084 AGCGGCTTAAAACGTCCAAATGTGCAGCATTTATCGGCGACGACAACATTATACACGGAG 7143
Qy 2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
Db 7144 TAGTATCTGACAAAGAAATGGCTGAGAGGTGTGCCACCTGGCTCAACATGGAGGTTAAGA 7203
Qy 2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139
Db 7204 TCATTGACGCAGTCATCGGCGAGAGACCACCTTACTTCTGCGGTGGATTCATCTTGCAAG 7263
Qy 2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199
Db 7264 ATTCGGTTACCTCCACAGCGTGTCGCGTGGCGGACCCCTTGAAAAGGCTGTTTAAGTTGG 7323
Qy 2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
Db 7324 GTAAACCGCTCCCAGCCGACGATGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 7383
Qy 2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319
Db 7384 AAACAAAGGCGTGGTTTAGAGTAGGTATAACAGACACCTTAGCAGTGGCCGTGGCAACTC 7443
Qy 2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379
Db 7444 GGTATGAGGTAGACAACATCACACCTGTCCTGCTGGCATTGAGAACTTTTGCCCAGAGCA 7503
Qy 2380 TTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGG 2439
Db 7504 AAAGAGCATTTCAAGCCATCAGAGGGGAAATAAAGCATCTCTACGGTGGTCCTAAATAGT 7563
Qy 2440 TGACGTAGTAGA----------CACGCACCTACCCACCGGCAGAATGTTTCCA 2482
Db 7564 CAGCATAGTACATTTCATCTGACTAATACCACAACACCACCACCATGAATAGAGGATTCT 7623
Qy 2483 TACCCTCAGCTGAACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCA 2542
Db 7624 TTAACATGCTCGGCCGCCGCCCCTTCCCAGCCCCCACTGCCATGTGGAGGCCGCGGAGAA 7683
Qy 2543 AACCCTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGAT 2602
Db 7684 GGAGGCAGGCGGCCCCGATGCCTGCCCGCAATGGGCTGGCTTCCCAAATCCAGCAACTGA 7743
Qy 2603 CTTAGGAGGTCGATAGTCAACTTGACTTTCAAACAAC--GATCACCTAATCCGCCGCCAG 2660
Db 7744 CCACAGCCGTCAGTGCCCTAGTCATTGGACAGGCAACTAGACCTCAAACCCCACGCCCAC 7803
Qy 2661 GTCCACCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGA 2720
Db 7804 GCCCGCCGCCGCGCCAGAAGAAGCAGGCGCCAAAGCAACCACCGAAGCCGAAGAAACCAA 7863
Qy 2721 AGAAGCAGCAAGCCAAGAGGACGAAAC---GCAAGCCTAAACCAGGGAAACGACAACGTA 2777
Db 7864 AAACACAGGAGAAGAAGAAGAAGCAACCTGCAAAACCCAAACCCGGAAAGAGACAGCGTA 7923
Qy 2778 TGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATC---ATGCTGAACGGCCAAGTGA 2834
Db 7924 TGGCACTTAAGTTGGAGGCCGACAGACTGTTCGACGTCAAAAATGAGGACGGAGATGTCA 7983
Qy 2835 ATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACTCCACGTTGAAGGAAAAA 2894
Db 7984 TCGGGCACGCACTGGCCATGGAAGGAAAGGTAATGAAACCACTCCACGTGAAAGGAACTA 8043
Qy 2895 TTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAGCATGTACGACTTGGAGT 2954
Db 8044 TTGACCACCCTGTGCTATCAAAGCTCAAATTCACCAAGTCGTCAGCATACGACATGGAGT 8103
Qy 2955 ACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTACACCAGCGACAAACCAC 3014
Db 8104 TCGCACAGTTGCCGGTCAACATGAGAAGTGAGGCGTTCACCTACACCAGTGAACACCCTG 8163
Qy 3015 CGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAATGGGAGATTTACCGTAC 3074
Db 8164 AAGGGTTCTACAACTGGCACCACGGAGCGGTGCAGTATAGTGGAGGCAGATTTACCATCC 8223
Qy 3075 CGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGACAACAGAGGCAGAG 3134
Db 8224 CCCGCGGAGTAGGAGGCAGAGGAGACAGTGGTCGTCCGATTATGGATAACTCAGGCCGGG 8283
Qy 3135 TTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGCGCTTTCAGTGGTCA 3194
Db 8284 TTGTCGCGATAGTCCTCGGAGGGGCTGATGAGGGAACAAGAACCGCCCTTTCGGTCGTCA 8343
Qy 3195 CTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTCTGAACCGTGGT--- 3251
Db 8344 CCTGGAATAGCAAAGGGAAGACAATCAAGACAACCCCGGAAGGGACAGAAGAGTGGTCTG 8403
Qy 3252 ------CACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGACAAAC 3305
Db 8404 CTGCACCACTGGTCACGGCCATGTGCTTGCTTGGAAACGTGAGCTTCCCATGCAATCGCC 8463
Qy 3306 CACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAACGTCG 3365
Db 8464 CGCCCACATGCTACACCCGCGAACCATCCAGAGCTCTCGACATCCTCGAAGAGAACGTGA 8523
Qy 3366 ACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAAT---GTCCATCACGCCGGC 3422
Db 8524 ACCACGAGGCCTACGACACCCTGCTCAACGCCATATTGCGGTGCGGATCGTCCGGCAGAA 8583
Qy 3423 CCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCCCGT 3482
Db 8584 GTAAAAGAAGCGTCACTGACGACTTTACCTTGACCAGCCCGTACTTGGGCACATGCTCGT 8643
Qy 3483 ATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACGAAT 3542
Db 8644 ACTGTCACCATACTGAACCGTGCTTTAGCCCGATTAAGATCGAGCAGGTCTGGGATGAAG 8703
Qy 3543 CTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAGGCA 3602
Db 8704 CGGACGACAACACCATACGCATACAGACTTCCGCCCAGTTTGGATACGACCAAAGCGGAG 8763
Qy 3603 CTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAAGGAAG 3662
Db 8764 CAGCAAGCTCAAATAAGTACCGCTACATGTCGCTCGAGCAGGATCATACTGTCAAAGAAG 8823
Qy 3663 ACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCACAAAG 3722
Db 8824 GCACCATGGATGACATCAAGATCAGCACCTCAGGACCGTGTAGAAGGCTTAGCTACAAAG 8883
Qy 3723 GGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCACGAGCG 3782
Db 8884 GATACTTTCTCCTCGCGAAGTGTCCTCCAGGGGACAGCGTAACGGTTAGCATAGCGAGTA 8943
Qy 3783 GAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTAGAG 3842
Db 8944 GCAACTCAGCAACGTCATGCACAATGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGG 9003
Qy 3843 AGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATCACT 3902
Db 9004 AAAAATATGACCTACCTCCCGTTCACGGTAAGAAGATTCCTTGCACAGTGTACGACCGTC 9063
Qy 3903 TGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAAGT 3962
Db 9064 TGAAAGAAACAACCGCCGGCTACATCACTATGCACAGGCCGGGACCGCATGCCTATACAT 9123
Qy 3963 CCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGTCA 4022
Db 9124 CCTATCTGGAGGAATCATCAGGGAAAGTTTACGCGAAGCCACCATCCGGGAAGAACATTA 9183
Qy 4023 CCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGATGA 4082
Db 9184 CGTACGAGTGCAAGTGCGGCGATTACAAGACCGGAACCGTTACGACCCGTACCGAAATCA 9243


Qy 4083 ACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGTCT 4142
Db 9244 CGGGCTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAACGAAGTGGGTCT 9303
Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9304 TCAACTCGCCGGACTCGATCAGACACGCCGACCACACGGCCCAAGGGAAATTGCATTTGC 9363
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9364 CTTTCAAGCTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAACGTAGTAC 9423
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9424 ACGGCTTTAAACACATCAGCCTCCAATTAGACACAGACCATCTGACATTGCTCACCACCA 9483
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9484 GGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCATCGGAAACACGGTTAGAA 9543
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9544 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGCAATCACGAACCAGTAAGGG 9603
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db 9604 TCTATGCCCAAGAGTCTGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9663
Qy 4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db 9664 ATTACTATCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTGCTGTGGCGA 9723
Qy 4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db 9724 TGATGATTGGCGTAACTGTTGCAGCATTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9783
Qy 4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db 9784 CGCCATATGCCCTGGCCCCAAATGCCGTGATTCCAACTTCGCTGGCACTTTTGTGCTGTG 9843
Qy 4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db 9844 TTAGGTCGGCTAATGCTGAAACATTCACCGAGACCATGAGTTACTTATGGTCGAACAGCC 9903
Qy 4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
Db 9904 AGCCGTTCTTCTGGGTCCAGCTGTGTATACCTCTGGCCGCTGTCGTCGTTCTAATGCGCT 9963
Qy 4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
Db 9964 GTTGCTCATGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10023
Qy 4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
Db 10024 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10083
Qy 4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db 10084 AAAGGGCAGGGTACGCCCCGCTCAATTTGGAGATTACTGTCATGTCCTCGGAGGTTTTGC 10143
Qy 4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db 10144 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCTAAAG 10203
Qy 5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db 10204 TCAGATGCTGCGGCTCCTTGGAATGTCAGCCCGCCGCTCACGCAGACTATACCTGCAAGG 10263
Qy 5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db 10264 TCTTTGGAGGGGTGTACCCCTTCATGTGGGGAGGAGCACAATGTTTTTGCGACAGTGAGA 10323

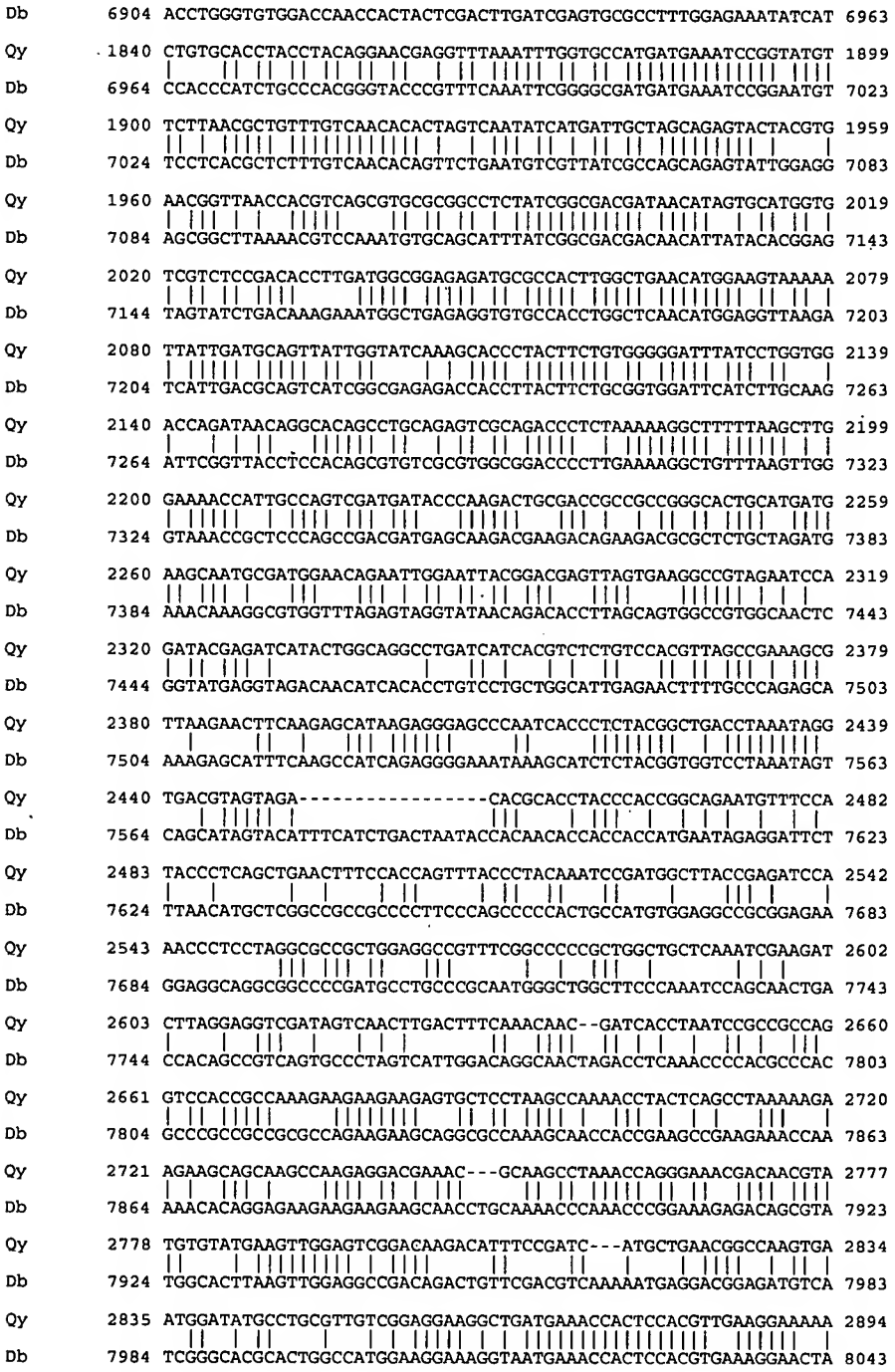

Qy 5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db 10324 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGTAGATTGCGCGACTGACCACGCGC 10383
Qy 5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db 10384 AGGCGATTAAGGTGCATACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10443
Qy 5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db 10444 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACCTGA 10503
Qy 5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db 10504 AAGTCATAGCTGGACCAATTTCAGCATTGTTTACACCATTCGATCACAAGGTCGTTATCA 10563
Qy 5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db 10564 ATCGCGGCCTGGTGTACAACTATGACTTTCCGGAATACGGAGCGATGAAACCAGGAGCGT 10623
Qy 5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db 10624 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAAGACCTCATCGCCAGCACAGACATTA 10683
Qy 5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db 10684 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCTGGAT 10743
Qy 5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db 10744 TCGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCCCCTTTTGGGTGCA 10803
Qy 5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db 10804 AGATTGCAGTCAATCCGCTTCGAGCGGTGGACTGCTCATACGGGAACATTCCCATTTCTA 10863
Qy 5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db 10864 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10923
Qy 5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db 10924 AATGTGATGTCAGTGAGTGCACTTATTCAGCGGACTTCGGAGGGATGGCTACCCTGCAGT 10983
Qy 5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db 10984 ATGTATCCGACCGCGAAGGACAATGCCCTGTACATTCGCATTCGAGCACAGCAACCCTCC 11043
Qy 5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db 11044 AAGAGTCGACAGTTCATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTCAGCACCGCGA 11103
Qy 5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db 11104 GCCCACAGGCGAACTTCATTGTATCGCTGTGTGGTAAGAAGACAACATGCAATGCAGAAT 11163
Qy 6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db 11164 GCAAACCACCAGCTGATCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTCCAAG 11223
Qy 6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db 11224 CCGCCATCTCAAAAACTTCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11283
Qy 6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
Db 11284 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11343
Qy 6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
Db 11344 GACCGCTACGCCCCAATGACCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11403
Qy 6242 GCATAATGCCACGCGCCGCT----------TGACACT 6268

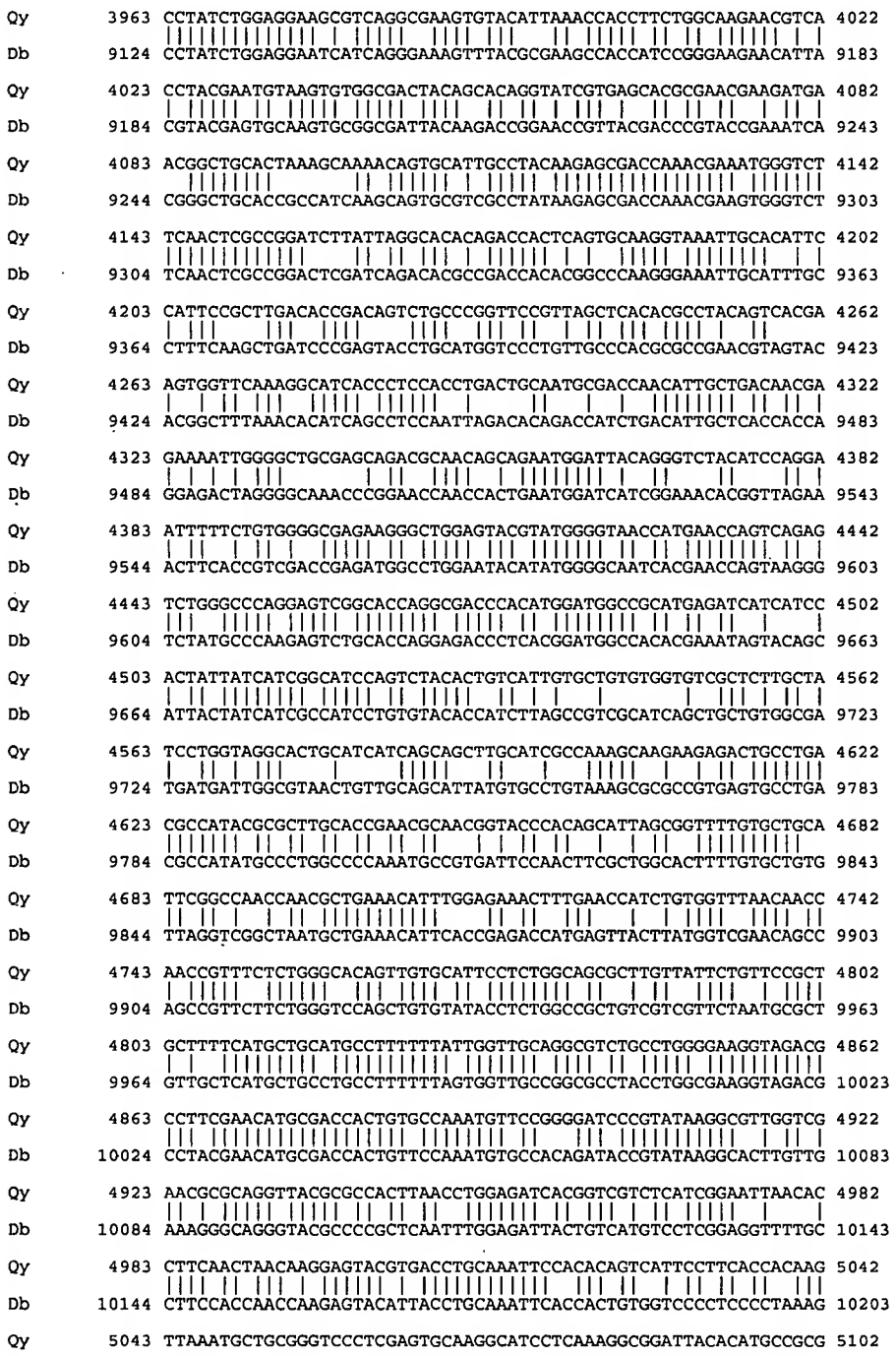
```
Db      11404 GCATAATGCATCAGGCTGGTATATTAGATCCCCGCTTACCGCGGGCAATATAGCAACACC 11463

Qy       6269 AAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGT 6318
Db      11464 AAAACTCGACGTATTTCCGAGGAAGCGCAGTGCATAATGCTGCGCAGTGT 11513
```

RESULT 7
US-09-102-248-1
; Sequence 1, Application US/09102248
; Patent No. 6008035
; GENERAL INFORMATION:
; APPLICANT: Johnston, Robert E.
; APPLICANT: Davis, Nancy L.
; APPLICANT: Simpson, Dennis A.
; TITLE OF INVENTION: System for the In Vivo Delivery and
; TITLE OF INVENTION: Expression of Heterologous Genes in the Bone Marrow
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson, P.A.
; STREET: 1211 East Morehead Street
; CITY: Charlotte
; STATE: No. 6008035th Carolina
; COUNTRY: USA
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/102,248
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/801,263
; FILING DATE: 19-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5470-147
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-420-2200
; TELEFAX: 919-881-3175
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11663 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 60..7559
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 7608..11342
US-09-102-248-1

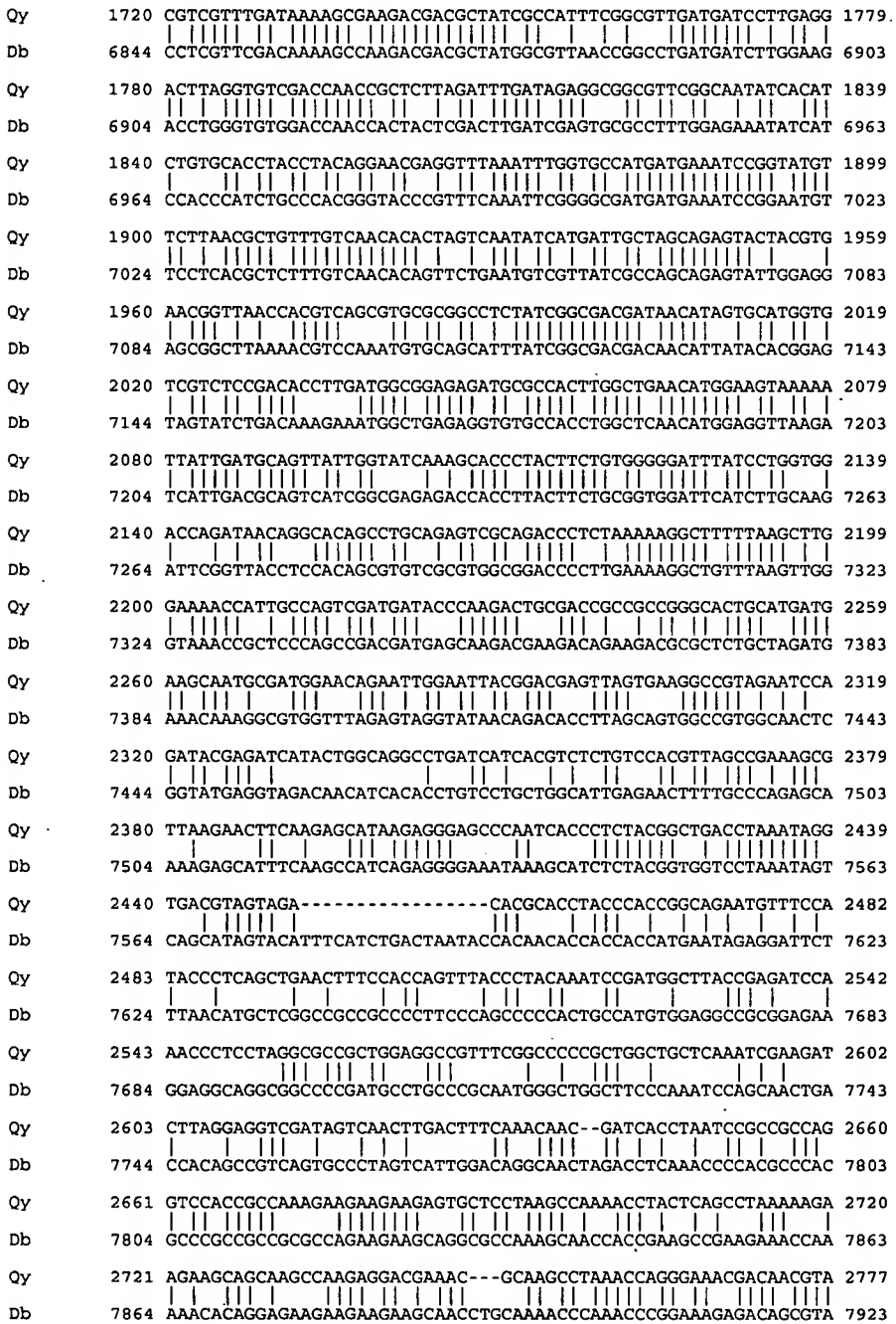
Query Match 36.9%; Score 2396; DB 3; Length 11663;
Best Local Similarity 64.6%; Pred. No. 0;
Matches 3751; Conservative 0; Mismatches 1985; Indels 74; Gaps 9;

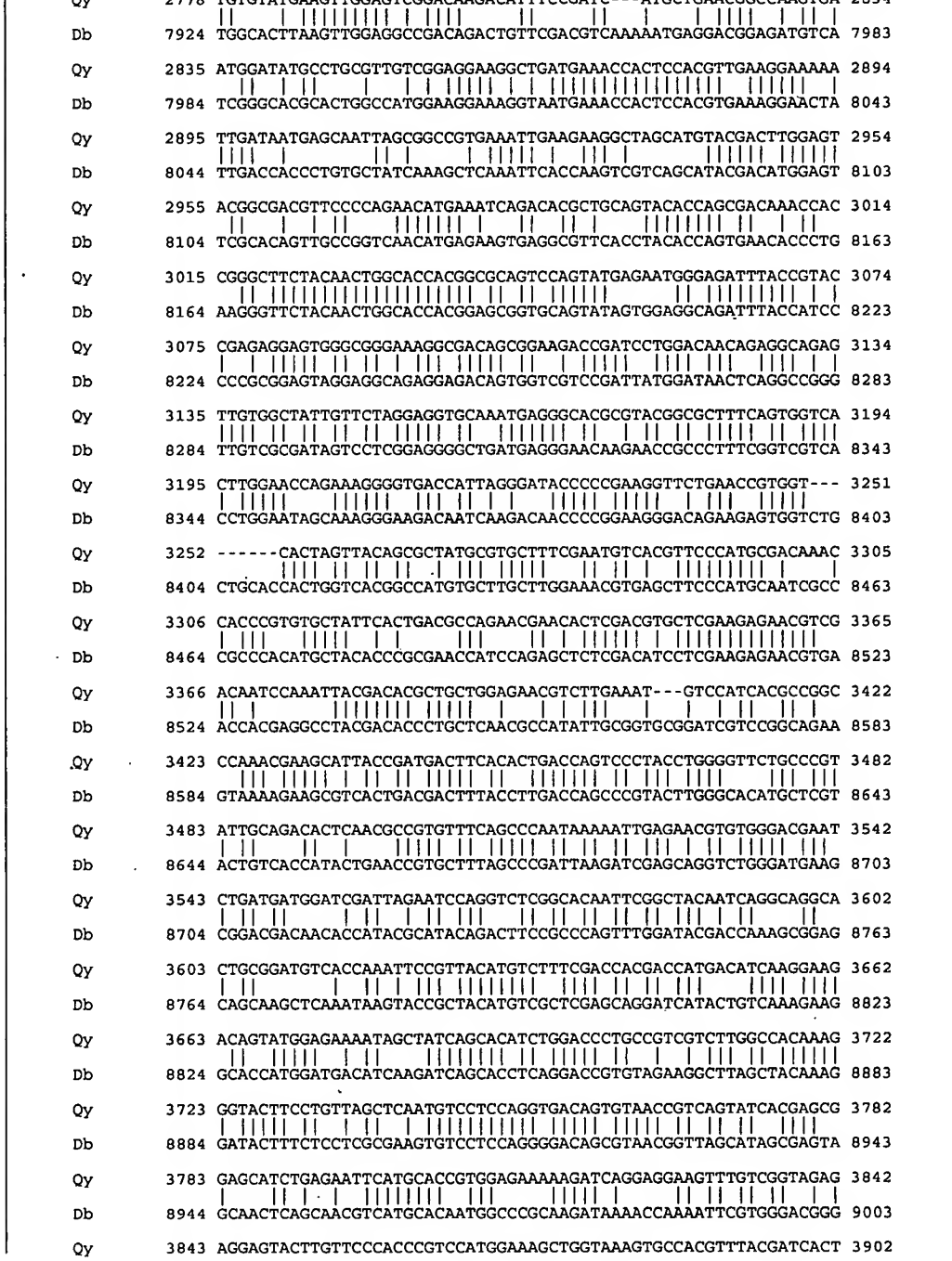
```
Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
Db       5707 ACTGTCTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5766

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
Db       5767 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTTCTGG 5826

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759

Db       5827 AAAGAATCTACGCCCCGGTGCTCGACACGTCGAAAGAGGAACAGCTCAAACTCAGGTACC 5886

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
Db       5887 AGATGATGCCCACCGAAGCCAACAAAAGCAGGTACCAGTCTCGAAAAGTAGAAAACCAGA 5946

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
Db       5947 AAGCCATAACCACTGAGCGACTGCTTTCAGGGCTACGACTGTATAACTCTGCCACAG--- 6003

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
Db       6004 ATCAGCCAGAATGCTATAAGATCACCTACCCGAAACCATCGTATTCCAGCAGTGTACCAG 6063

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
Db       6064 CGAACTACTCTGACCCAAAGTTTGCTGTAGCTGTTTGTAACAACTATCTGCATGAGAATT 6123

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
Db       6124 ACCCGACGGTAGCATCTTATCAGATCACCGACGAGTACGATGCTTACTTGGATATGGTAG 6183

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
Db       6184 ACGGGACAGTCGCTTGCCTAGATACTGCAACTTTTTGCCCCGCCAAGCTTAGAAGTTACC 6243

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
Db       6244 CGAAAAGACACGAGTATAGAGCCCCAAACATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6303

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
Db       6304 ACACGTTGCAAAACGTGCTCATTGCCGCGACTAAAAGAAACTGCAACGTCACACAAATGC 6363

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
Db       6364 GTGAACTGCCAACACTGGACTCAGCGACATTCAACGTTGAATGCTTTCGAAAATATGCAT 6423

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
Db       6424 GCAATGACGAGTATTGGGAGGAGTTTGCCCGAAAGCCAATTAGGATCACTACTGAGTTCG 6483

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
Db       6484 TTACCGCATACGTGGCCAGACTGAAAGGCCCTAAGGCCGCCGCACTGTTCGCAAAGACGC 6543

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
Db       6544 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6603

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
Db       6604 ACGTGAAAGTTACACCTGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6663

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
Db       6664 AAGCCGCAGAACCCCTGGCGACCGCTTACCTATGCGGGATCCACCGGGAGTTAGTGCGCA 6723

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
Db       6724 GGCTTACAGCCGTTTTGCTACCCAACATTCACACGCTCTTTGACATGTCGGCGGAGGACT 6783

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
Db       6784 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGTGACCCGGTACTGGAGACGGATATCG 6843

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
Db       6844 CCTCGTTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGCCTGATGATCTTGGAAG 6903

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
```

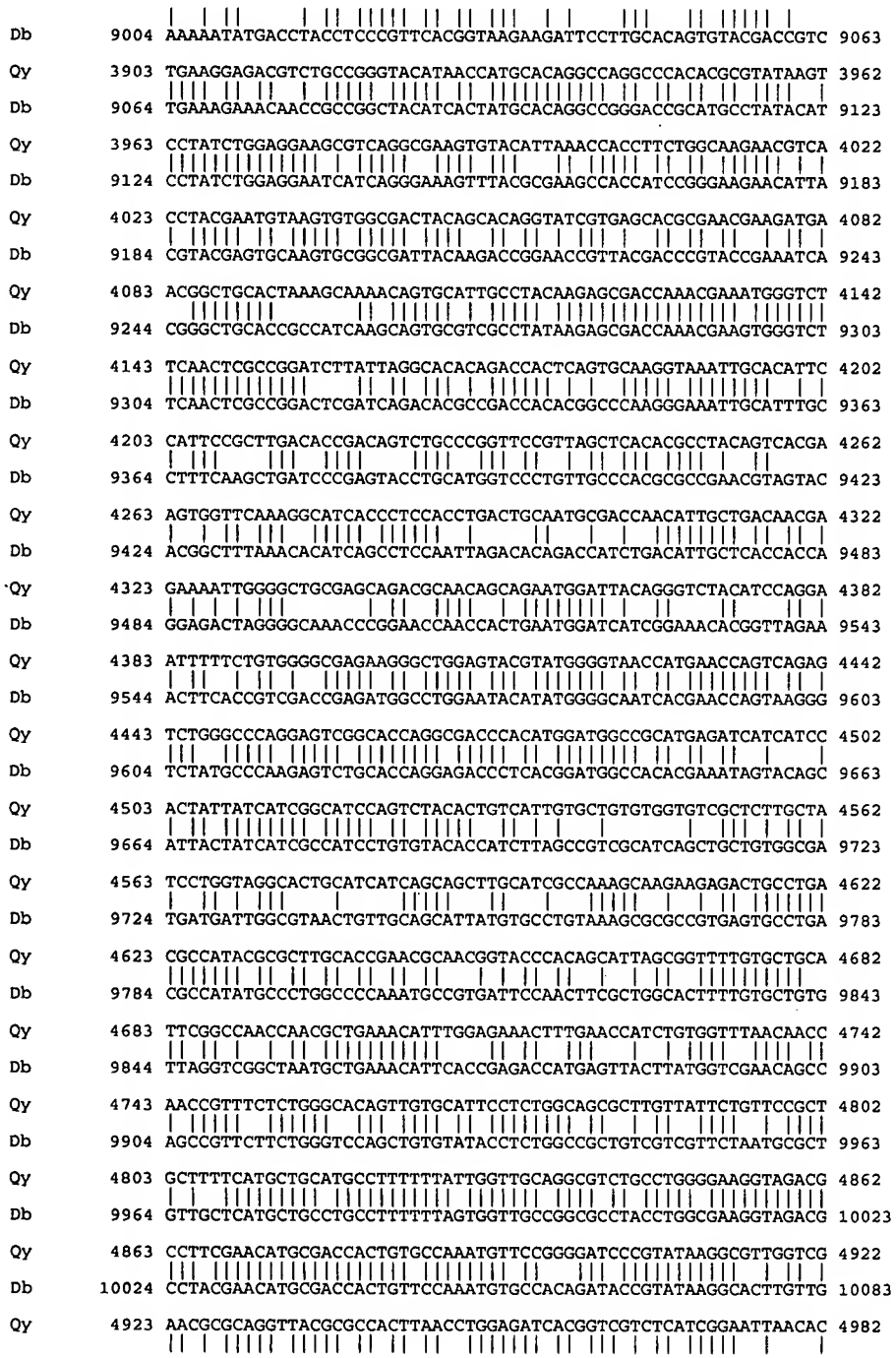

Db 6904 ACCTGGGTGTGGACCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 6963
Qy 1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
Db 6964 CCACCCATCTGCCCACGGGTACCCGTTTCAAATTCGGGGCGATGATGAAATCCGGAATGT 7023
Qy 1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
Db 7024 TCCTCACGCTCTTTGTCAACACAGTTCTGAATGTCGTTATCGCCAGCAGAGTATTGGAGG 7083
Qy 1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
Db 7084 AGCGGCTTAAAACGTCCAAATGTGCAGCATTTATCGGCGACGACAACATTATACACGGAG 7143
Qy 2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
Db 7144 TAGTATCTGACAAAGAAATGGCTGAGAGGTGTGCCACCTGGCTCAACATGGAGGTTAAGA 7203
Qy 2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139
Db 7204 TCATTGACGCAGTCATCGGCGAGAGACCACCTTACTTCTGCGGTGGATTCATCTTGCAAG 7263
Qy 2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199
Db 7264 ATTCGGTTACCTCCACAGCGTGTCGCGTGGCGGACCCCTTGAAAAGGCTGTTTAAGTTGG 7323
Qy 2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
Db 7324 GTAAACCGCTCCCAGCCGACGATGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 7383
Qy 2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319
Db 7384 AAACAAAGGCGTGGTTTAGAGTAGGTATAACAGACACCTTAGCAGTGGCCGTGGCAACTC 7443
Qy 2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379
Db 7444 GGTATGAGGTAGACAACATCACACCTGTCCTGCTGGCATTGAGAACTTTTGCCCAGAGCA 7503
Qy 2380 TTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGG 2439
Db 7504 AAAGAGCATTTCAAGCCATCAGAGGGGAAATAAAGCATCTCTACGGTGGTCCTAAATAGT 7563
Qy 2440 TGACGTAGTAGA----------CACGCACCTACCCACCGGCAGAATGTTTCCA 2482
Db 7564 CAGCATAGTACATTTCATCTGACTAATACCACAACACCACCACCATGAATAGAGGATTCT 7623
Qy 2483 TACCCTCAGCTGAACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCA 2542
Db 7624 TTAACATGCTCGGCCGCCGCCCCTTCCCAGCCCCCACTGCCATGTGGAGGCCGCGGAGAA 7683
Qy 2543 AACCCTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGAT 2602
Db 7684 GGAGGCAGGCGGCCCCGATGCCTGCCCGCAATGGGCTGGCTTCCCAAATCCAGCAACTGA 7743
Qy 2603 CTTAGGAGGTCGATAGTCAACTTGACTTTCAAACAAC--GATCACCTAATCCGCCGCCAG 2660
Db 7744 CCACAGCCGTCAGTGCCCTAGTCATTGGACAGGCAACTAGACCTCAAACCCCACGCCCAC 7803
Qy 2661 GTCCACCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGA 2720
Db 7804 GCCCGCCGCCGCGCCAGAAGAAGCAGGCGCCAAAGCAACCACCGAAGCCGAAGAAACCAA 7863
Qy 2721 AGAAGCAGCAAGCCAAGAGGACGAAAC---GCAAGCCTAAACCAGGGAAACGACAACGTA 2777
Db 7864 AAACACAGGAGAAGAAGAAGAAGCAACCTGCAAAACCCAAACCCGGAAAGAGACAGCGTA 7923
Qy 2778 TGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATC---ATGCTGAACGGCCAAGTGA 2834
Db 7924 TGGCACTTAAGTTGGAGGCCGACAGACTGTTCGACGTCAAAAATGAGGACGGAGATGTCA 7983
Qy 2835 ATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACTCCACGTTGAAGGAAAAA 2894
Db 7984 TCGGGCACGCACTGGCCATGGAAGGAAAGGTAATGAAACCACTCCACGTGAAAGGAACTA 8043
Qy 2895 TTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAGCATGTACGACTTGGAGT 2954
Db 8044 TTGACCACCCTGTGCTATCAAAGCTCAAATTCACCAAGTCGTCAGCATACGACATGGAGT 8103
Qy 2955 ACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTACACCAGCGACAAACCAC 3014
Db 8104 TCGCACAGTTGCCGGTCAACATGAGAAGTGAGGCGTTCACCTACACCAGTGAACACCCTG 8163
Qy 3015 CGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAATGGGAGATTTACCGTAC 3074
Db 8164 AAGGGTTCTACAACTGGCACCACGGAGCGGTGCAGTATAGTGGAGGCAGATTTACCATCC 8223
Qy 3075 CGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGACAACAGAGGCAGAG 3134
Db 8224 CCCGCGGAGTAGGAGGCAGAGGAGACAGTGGTCGTCCGATTATGGATAACTCAGGCCGGG 8283
Qy 3135 TTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGCGCTTTCAGTGGTCA 3194
Db 8284 TTGTCGCGATAGTCCTCGGAGGGGCTGATGAGGGAACAAGAACCGCCCTTTCGGTCGTCA 8343
Qy 3195 CTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTCTGAACCGTGGT--- 3251
Db 8344 CCTGGAATAGCAAAGGGAAGACAATCAAGACAACCCCGGAAGGGACAGAAGAGTGGTCTG 8403
Qy 3252 ------CACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGACAAAC 3305
Db 8404 CTGCACCACTGGTCACGGCCATGTGCTTGCTTGGAAACGTGAGCTTCCCATGCAATCGCC 8463
Qy 3306 CACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAACGTCG 3365
Db 8464 CGCCCACATGCTACACCCGCGAACCATCCAGAGCTCTCGACATCCTCGAAGAGAACGTGA 8523
Qy 3366 ACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAAT---GTCCATCACGCCGGC 3422
Db 8524 ACCACGAGGCCTACGACACCCTGCTCAACGCCATATTGCGGTGCGGATCGTCCGGCAGAA 8583
Qy 3423 CCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCCCGT 3482
Db 8584 GTAAAAGAAGCGTCACTGACGACTTTACCTTGACCAGCCCGTACTTGGGCACATGCTCGT 8643
Qy 3483 ATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACGAAT 3542
Db 8644 ACTGTCACCATACTGAACCGTGCTTTAGCCCGATTAAGATCGAGCAGGTCTGGGATGAAG 8703
Qy 3543 CTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAGGCA 3602
Db 8704 CGGACGACAACACCATACGCATACAGACTTCCGCCCAGTTTGGATACGACCAAAGCGGAG 8763
Qy 3603 CTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAAGGAAG 3662
Db 8764 CAGCAAGCTCAAATAAGTACCGCTACATGTCGCTCGAGCAGGATCATACTGTCAAAGAAG 8823
Qy 3663 ACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCACAAAG 3722
Db 8824 GCACCATGGATGACATCAAGATCAGCACCTCAGGACCGTGTAGAAGGCTTAGCTACAAAG 8883
Qy 3723 GGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCACGAGCG 3782
Db 8884 GATACTTTCTCCTCGCGAAGTGTCCTCCAGGGGACAGCGTAACGGTTAGCATAGCGAGTA 8943
Qy 3783 GAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTAGAG 3842
Db 8944 GCAACTCAGCAACGTCATGCACAATGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGG 9003
Qy 3843 AGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATCACT 3902
Db 9004 AAAAATATGACCTACCTCCCGTTCACGGTAAGAAGATTCCTTGCACAGTGTACGACCGTC 9063
Qy 3903 TGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAAGT 3962
Db 9064 TGAAAGAAACAACCGCCGGCTACATCACTATGCACAGGCCGGGACCGCATGCCTATACAT 9123


Qy 3963 CCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGTCA 4022
Db 9124 CCTATCTGGAGGAATCATCAGGGAAAGTTTACGCGAAGCCACCATCCGGGAAGAACATTA 9183
Qy 4023 CCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGATGA 4082
Db 9184 CGTACGAGTGCAAGTGCGGCGATTACAAGACCGGAACCGTTACGACCCGTACCGAAATCA 9243
Qy 4083 ACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGTCT 4142
Db 9244 CGGGCTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAACGAAGTGGGTCT 9303
Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9304 TCAACTCGCCGGACTCGATCAGACACGCCGACCACACGGCCCAAGGGAAATTGCATTTGC 9363
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9364 CTTTCAAGCTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAACGTAGTAC 9423
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9424 ACGGCTTTAAACACATCAGCCTCCAATTAGACACAGACCATCTGACATTGCTCACCACCA 9483
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9484 GGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCATCGGAAACACGGTTAGAA 9543
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9544 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGCAATCACGAACCAGTAAGGG 9603
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db 9604 TCTATGCCCAAGAGTCTGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9663
Qy 4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db 9664 ATTACTATCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTGCTGTGGCGA 9723
Qy 4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db 9724 TGATGATTGGCGTAACTGTTGCAGCATTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9783
Qy 4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db 9784 CGCCATATGCCCTGGCCCCAAATGCCGTGATTCCAACTTCGCTGGCACTTTTGTGCTGTG 9843
Qy 4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db 9844 TTAGGTCGGCTAATGCTGAAACATTCACCGAGACCATGAGTTACTTATGGTCGAACAGCC 9903
Qy 4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
Db 9904 AGCCGTTCTTCTGGGTCCAGCTGTGTATACCTCTGGCCGCTGTCGTCGTTCTAATGCGCT 9963
Qy 4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
Db 9964 GTTGCTCATGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10023
Qy 4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
Db 10024 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10083
Qy 4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db 10084 AAAGGGCAGGGTACGCCCCGCTCAATTTGGAGATTACTGTCATGTCCTCGGAGGTTTTGC 10143
Qy 4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db 10144 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCTAAAG 10203
Qy 5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db 10204 TCAGATGCTGCGGCTCCTTGGAATGTCAGCCCGCCGCTCACGCAGACTATACCTGCAAGG 10263
Qy 5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db 10264 TCTTTGGAGGGGTGTACCCCTTCATGTGGGGAGGAGCACAATGTTTTTGCGACAGTGAGA 10323
Qy 5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db 10324 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGTAGATTGCGCGACTGACCACGCGC 10383
Qy 5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db 10384 AGGCGATTAAGGTGCATACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10443
Qy 5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db 10444 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACCTGA 10503
Qy 5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db 10504 AAGTCATAGCTGGACCAATTTCAGCATTGTTTACACCATTCGATCACAAGGTCGTTATCA 10563
Qy 5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db 10564 ATCGCGGCCTGGTGTACAACTATGACTTTCCGGAATACGGAGCGATGAAACCAGGAGCGT 10623
Qy 5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db 10624 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAAGACCTCATCGCCAGCACAGACATTA 10683
Qy 5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db 10684 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCTGGAT 10743
Qy 5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db 10744 TCGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCCCCTTTTGGGTGCA 10803
Qy 5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db 10804 AGATTGCAGTCAATCCGCTTCGAGCGGTGGACTGCTCATACGGGAACATTCCCATTTCTA 10863
Qy 5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db 10864 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10923
Qy 5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db 10924 AATGTGATGTCAGTGAGTGCACTTATTCAGCGGACTTCGGAGGGATGGCTACCCTGCAGT 10983
Qy 5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db 10984 ATGTATCCGACCGCGAAGGACAATGCCCTGTACATTCGCATTCGAGCACAGCAACCCTCC 11043
Qy 5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db 11044 AAGAGTCGACAGTTCATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTCAGCACCGCGA 11103
Qy 5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db 11104 GCCCACAGGCGAACTTCATTGTATCGCTGTGTGGTAAGAAGACAACATGCAATGCAGAAT 11163
Qy 6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db 11164 GCAAACCACCAGCTGATCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTCCAAG 11223
Qy 6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db 11224 CCGCCATCTCAAAAACTTCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11283
Qy 6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182

```
Db      11284 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11343

Qy       6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241

Db      11344 GACCGCTACGCCCCAATGACCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11403

Qy       6242 GCATAATGCCACGCGCCGCT----------------------------------TGACACT 6268

Db      11404 GCATAATGCATCAGGCTGGTATATTAGATCCCCGCTTACCGCGGGCAATATAGCAACACC 11463

Qy       6269 AAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGT 6318

Db      11464 AAAACTCGACGTATTTCCGAGGAAGCGCAGTGCATAATGCTGCGCAGTGT 11513


RESULT 8
US-09-367-764-1
; Sequence 1, Application US/09367764
; Patent No. 6583121
;  GENERAL INFORMATION:
;    APPLICANT:  Johnston, Robert E.
;    APPLICANT:  Davis, Nancy L.
;    APPLICANT:  Simpson, Dennis A.
;    TITLE OF INVENTION:  System for the In Vivo Delivery and
;    TITLE OF INVENTION:  Expression of Heterologous Genes in the Bone Marrow
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson, P.A.
;      STREET:  1211 East Morehead Street
;      CITY:  Charlotte
;      STATE:  No. 6583121th Carolina
;      COUNTRY:  USA
;      ZIP:  28234
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/367,764
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/801,263
;      FILING DATE:  19-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sibley, Kenneth D.
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5470-147
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-420-2200
;      TELEFAX:  919-881-3175
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11663 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  60..7559
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  7608..11342
US-09-367-764-1

  Query Match             36.9%;  Score 2396;  DB 4;  Length 11663;
  Best Local Similarity   64.6%;  Pred. No. 0;
  Matches 3751;  Conservative    0;  Mismatches 1985;  Indels   74;  Gaps    9;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639

Db       5707 ACTGTCTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5766

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699

Db       5767 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTTCTGG 5826

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759

Db       5827 AAAGAATCTACGCCCCGGTGCTCGACACGTCGAAAGAGGAACAGCTCAAACTCAGGTACC 5886

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819

Db       5887 AGATGATGCCCACCGAAGCCAACAAAAGCAGGTACCAGTCTCGAAAAGTAGAAAACCAGA 5946

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879

Db       5947 AAGCCATAACCACTGAGCGACTGCTTTCAGGGCTACGACTGTATAACTCTGCCACAG--- 6003

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939

Db       6004 ATCAGCCAGAATGCTATAAGATCACCTACCCGAAACCATCGTATTCCAGCAGTGTACCAG 6063

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999

Db       6064 CGAACTACTCTGACCCAAAGTTTGCTGTAGCTGTTTGTAACAACTATCTGCATGAGAATT 6123

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059

Db       6124 ACCCGACGGTAGCATCTTATCAGATCACCGACGAGTACGATGCTTACTTGGATATGGTAG 6183

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119

Db       6184 ACGGGACAGTCGCTTGCCTAGATACTGCAACTTTTTGCCCCGCCAAGCTTAGAAGTTACC 6243

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179

Db       6244 CGAAAAGACACGAGTATAGAGCCCCAAACATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6303

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239

Db       6304 ACACGTTGCAAAACGTGCTCATTGCCGCGACTAAAAGAAACTGCAACGTCACACAAATGC 6363

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299

Db       6364 GTGAACTGCCAACACTGGACTCAGCGACATTCAACGTTGAATGCTTTCGAAAATATGCAT 6423

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359

Db       6424 GCAATGACGAGTATTGGGAGGAGTTTGCCCGAAAGCCAATTAGGATCACTACTGAGTTCG 6483

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419

Db       6484 TTACCGCATACGTGGCCAGACTGAAAGGCCCTAAGGCCGCCGCACTGTTCGCAAAGACGC 6543

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479

Db       6544 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6603

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539

Db       6604 ACGTGAAAGTTACACCTGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6663

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599

Db       6664 AAGCCGCAGAACCCCTGGCGACCGCTTACCTATGCGGGATCCACCGGGAGTTAGTGCGCA 6723

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659

Db       6724 GGCTTACAGCCGTTTTGCTACCCAACATTCACACGCTCTTTGACATGTCGGCGGAGGACT 6783

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719

Db       6784 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGTGACCCGGTACTGGAGACGGATATCG 6843
```





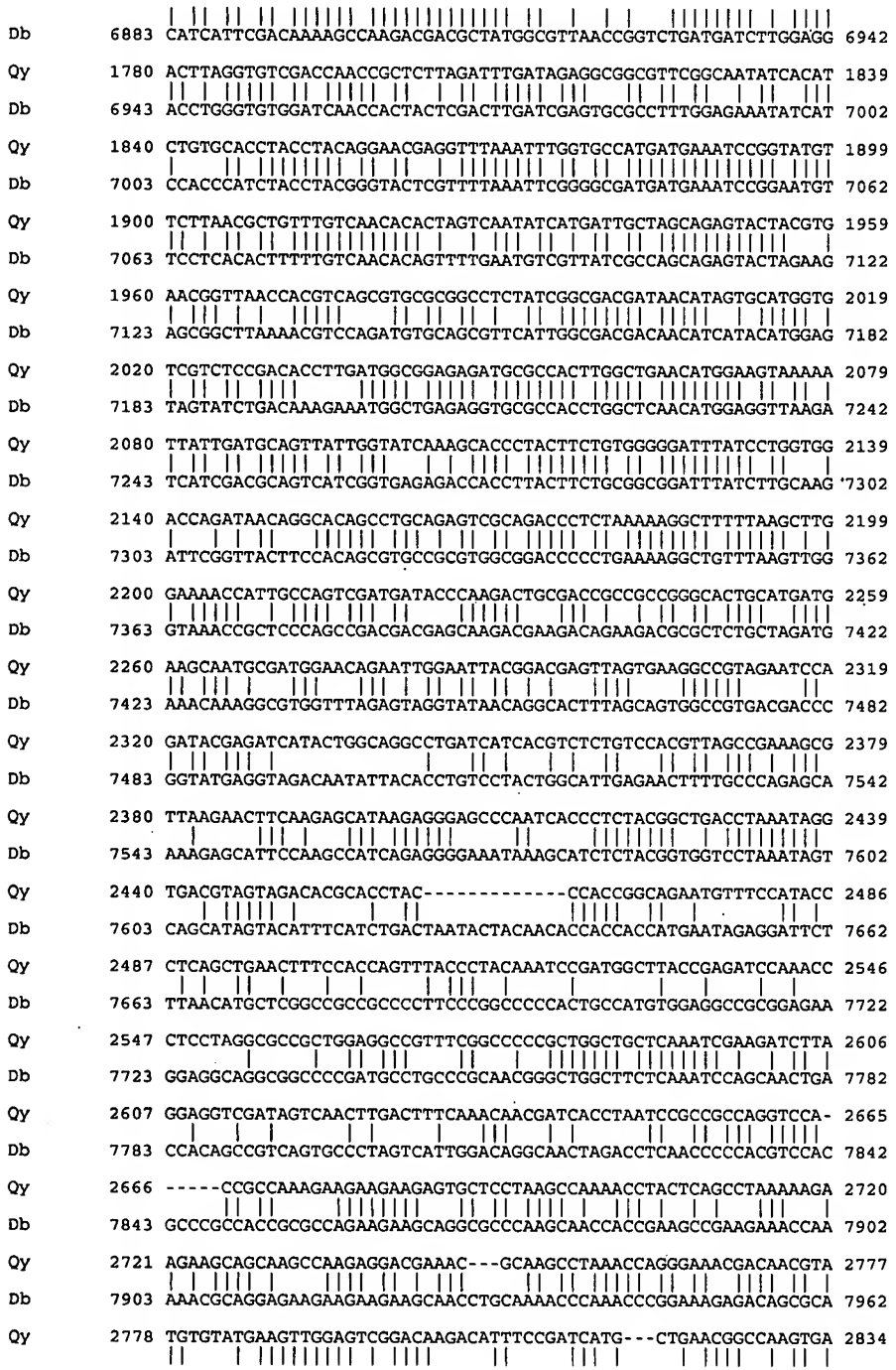

Db 9004 AAAAATATGACCTACCTCCCGTTCACGGTAAGAAGATTCCTTGCACAGTGTACGACCGTC 9063
Qy 3903 TGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAAGT 3962
Db 9064 TGAAAGAAACAACCGCCGGCTACATCACTATGCACAGGCCGGGACCGCATGCCTATACAT 9123
Qy 3963 CCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGTCA 4022
Db 9124 CCTATCTGGAGGAATCATCAGGGAAAGTTTACGCGAAGCCACCATCCGGGAAGAACATTA 9183
Qy 4023 CCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGATGA 4082
Db 9184 CGTACGAGTGCAAGTGCGGCGATTACAAGACCGGAACCGTTACGACCCGTACCGAAATCA 9243
Qy 4083 ACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGTCT 4142
Db 9244 CGGGCTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAACGAAGTGGGTCT 9303
Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9304 TCAACTCGCCGGACTCGATCAGACACGCCGACCACACGGCCCAAGGGAAATTGCATTTGC 9363
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9364 CTTTCAAGCTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAACGTAGTAC 9423
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9424 ACGGCTTTAAACACATCAGCCTCCAATTAGACACAGACCATCTGACATTGCTCACCACCA 9483
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9484 GGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCATCGGAAACACGGTTAGAA 9543
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9544 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGCAATCACGAACCAGTAAGGG 9603
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db 9604 TCTATGCCCAAGAGTCTGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9663
Qy 4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db 9664 ATTACTATCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTGCTGTGGCGA 9723
Qy 4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db 9724 TGATGATTGGCGTAACTGTTGCAGCATTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9783
Qy 4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db 9784 CGCCATATGCCCTGGCCCCAAATGCCGTGATTCCAACTTCGCTGGCACTTTTGTGCTGTG 9843
Qy 4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db 9844 TTAGGTCGGCTAATGCTGAAACATTCACCGAGACCATGAGTTACTTATGGTCGAACAGCC 9903
Qy 4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
Db 9904 AGCCGTTCTTCTGGGTCCAGCTGTGTATACCTCTGGCCGCTGTCGTCGTTCTAATGCGCT 9963
Qy 4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
Db 9964 GTTGCTCATGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10023
Qy 4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
Db 10024 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10083
Qy 4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db 10084 AAAGGGCAGGGTACGCCCCGCTCAATTTGGAGATTACTGTCATGTCCTCGGAGGTTTTGC 10143
Qy 4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db 10144 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCTAAAG 10203
Qy 5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db 10204 TCAGATGCTGCGGCTCCTTGGAATGTCAGCCCGCCGCTCACGCAGACTATACCTGCAAGG 10263
Qy 5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db 10264 TCTTTGGAGGGGTGTACCCCTTCATGTGGGGAGGAGCACAATGTTTTTGCGACAGTGAGA 10323
Qy 5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db 10324 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGTAGATTGCGCGACTGACCACGCGC 10383
Qy 5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db 10384 AGGCGATTAAGGTGCATACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10443
Qy 5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db 10444 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACCTGA 10503
Qy 5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db 10504 AAGTCATAGCTGGACCAATTTCAGCATTGTTTACACCATTCGATCACAAGGTCGTTATCA 10563
Qy 5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db 10564 ATCGCGGCCTGGTGTACAACTATGACTTTCCGGAATACGGAGCGATGAAACCAGGAGCGT 10623
Qy 5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db 10624 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAAGACCTCATCGCCAGCACAGACATTA 10683
Qy 5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db 10684 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCTGGAT 10743
Qy 5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db 10744 TCGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCCCCCTTTTGGGTGCA 10803
Qy 5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db 10804 AGATTGCAGTCAATCCGCTTCGAGCGGTGGACTGCTCATACGGGAACATTCCCATTTCTA 10863
Qy 5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db 10864 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10923
Qy 5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db 10924 AATGTGATGTCAGTGAGTGCACTTATTCAGCGGACTTCGGAGGGATGGCTACCCTGCAGT 10983
Qy 5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db 10984 ATGTATCCGACCGCGAAGGACAATGCCCTGTACATTCGCATTCGAGCACAGCAACCCTCC 11043
Qy 5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db 11044 AAGAGTCGACAGTTCATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTCAGCACCGCGA 11103
Qy 5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db 11104 GCCCACAGGCGAACTTCATTGTATCGCTGTGTGGTAAGAAGACAACATGCAATGCAGAAT 11163
Qy 6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db 11164 GCAAACCACCAGCTGATCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTCCAAG 11223

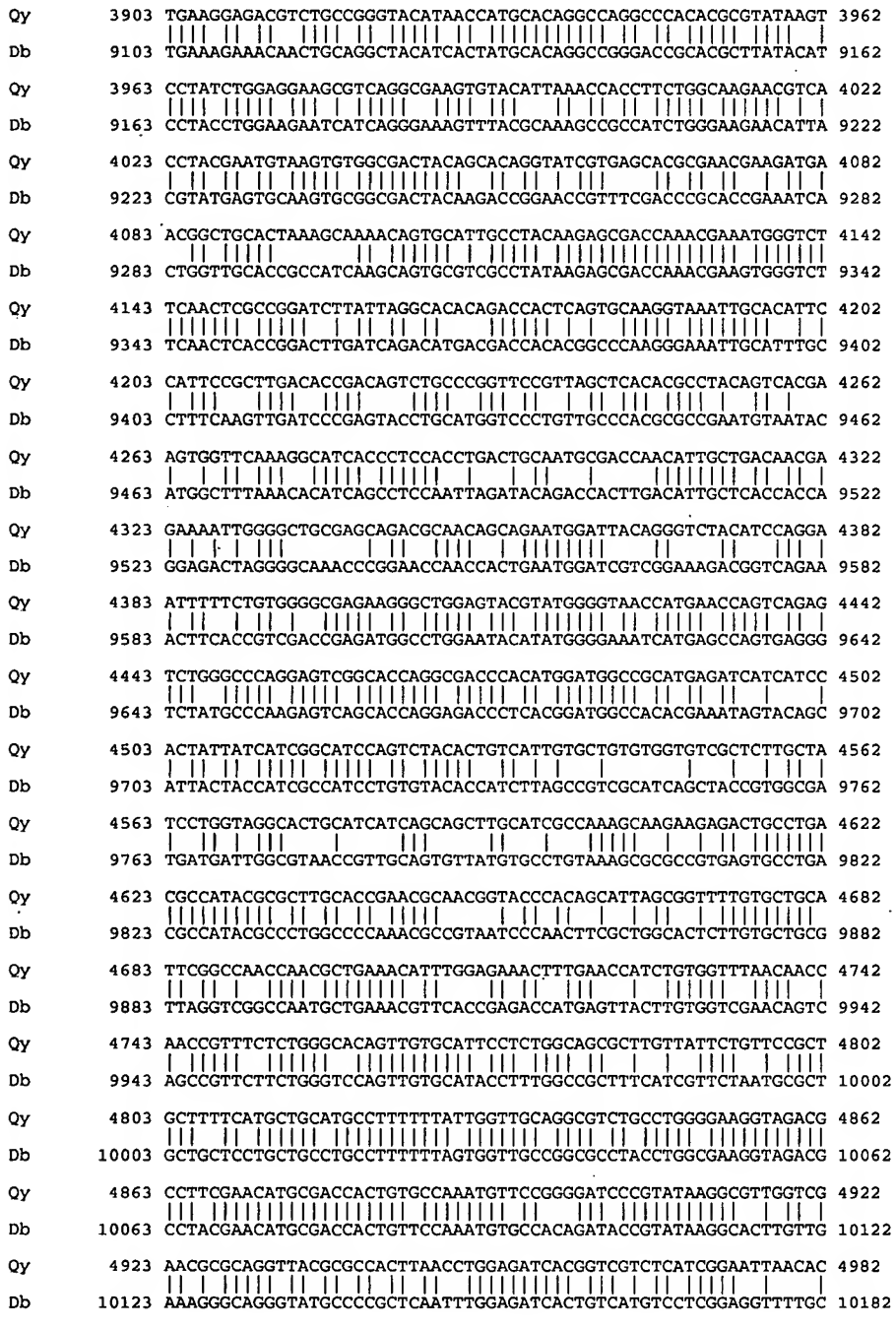
```
Qy       6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db      11224 CCGCCATCTCAAAAACTTCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11283

Qy       6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
Db      11284 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11343

Qy       6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
Db      11344 GACCGCTACGCCCCAATGACCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11403

Qy       6242 GCATAATGCCACGCGCCGCT--------------------------------TGACACT 6268
Db      11404 GCATAATGCATCAGGCTGGTATATTAGATCCCCGCTTACCGCGGGCAATATAGCAACACC 11463

Qy       6269 AAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGT 6318
Db      11464 AAAACTCGACGTATTTCCGAGGAAGCGCAGTGCATAATGCTGCGCAGTGT 11513
```

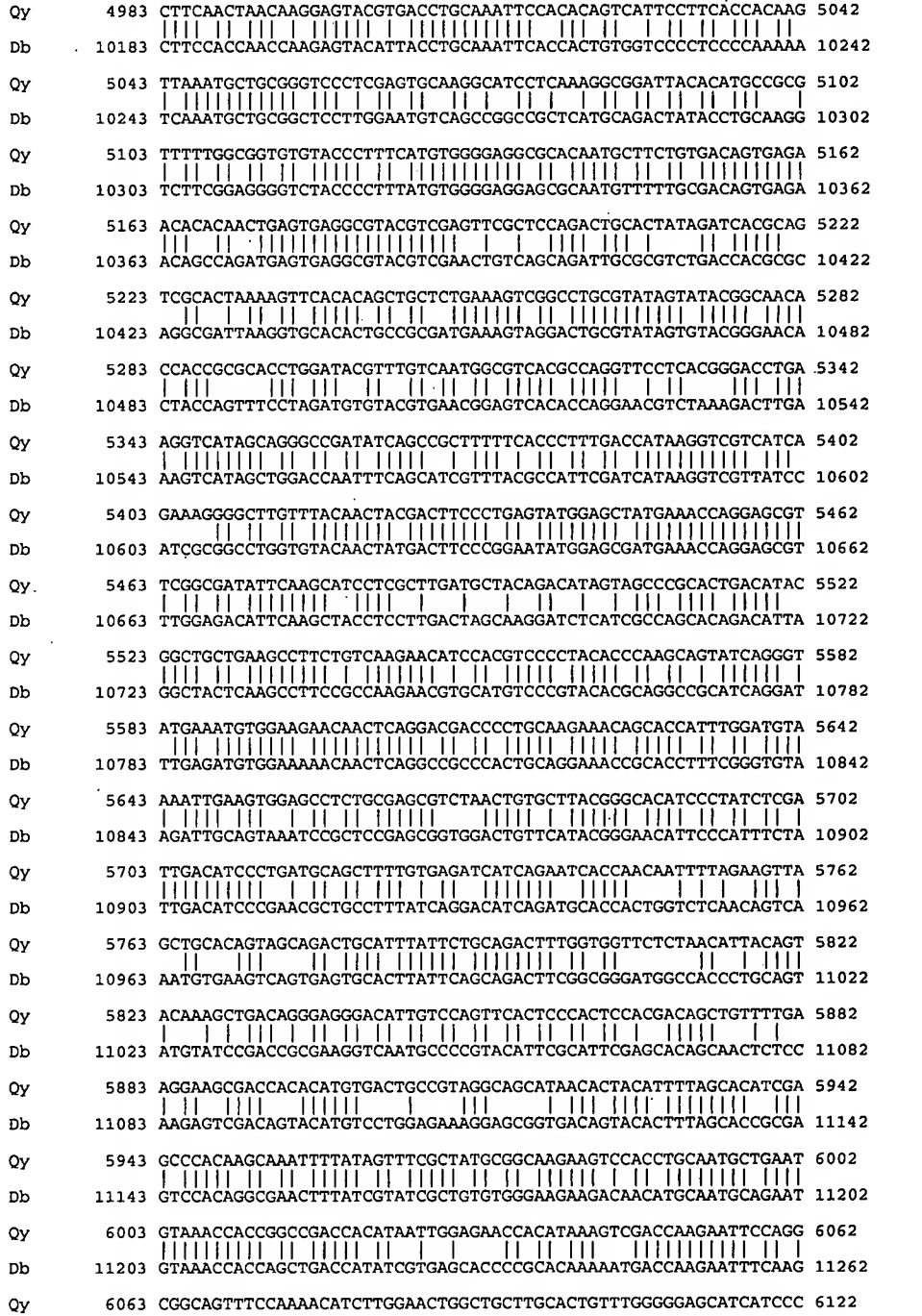
RESULT 9
US-08-801-263A-8
; Sequence 8, Application US/08801263A
; Patent No. 5811407
; GENERAL INFORMATION:
; APPLICANT: Johnston, Robert E.
; APPLICANT: Davis, Nancy L.
; APPLICANT: Simpson, Dennis A.
; TITLE OF INVENTION: System for the In Vivo Delivery and
; TITLE OF INVENTION: Expression of Heterologous Genes in the Bone Marrow
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson, P.A.
; STREET: 1211 East Morehead Street
; CITY: Charlotte
; STATE: No. 5811407th Carolina
; COUNTRY: USA
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,263A
; FILING DATE: 19-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5470-147
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-420-2200
; TELEFAX: 919-881-3175
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11703 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-801-263A-8

Query Match 36.8%; Score 2388.4; DB 1; Length 11703;
Best Local Similarity 64.7%; Pred. No. 0;
Matches 3698; Conservative 0; Mismatches 1981; Indels 41; Gaps 8;

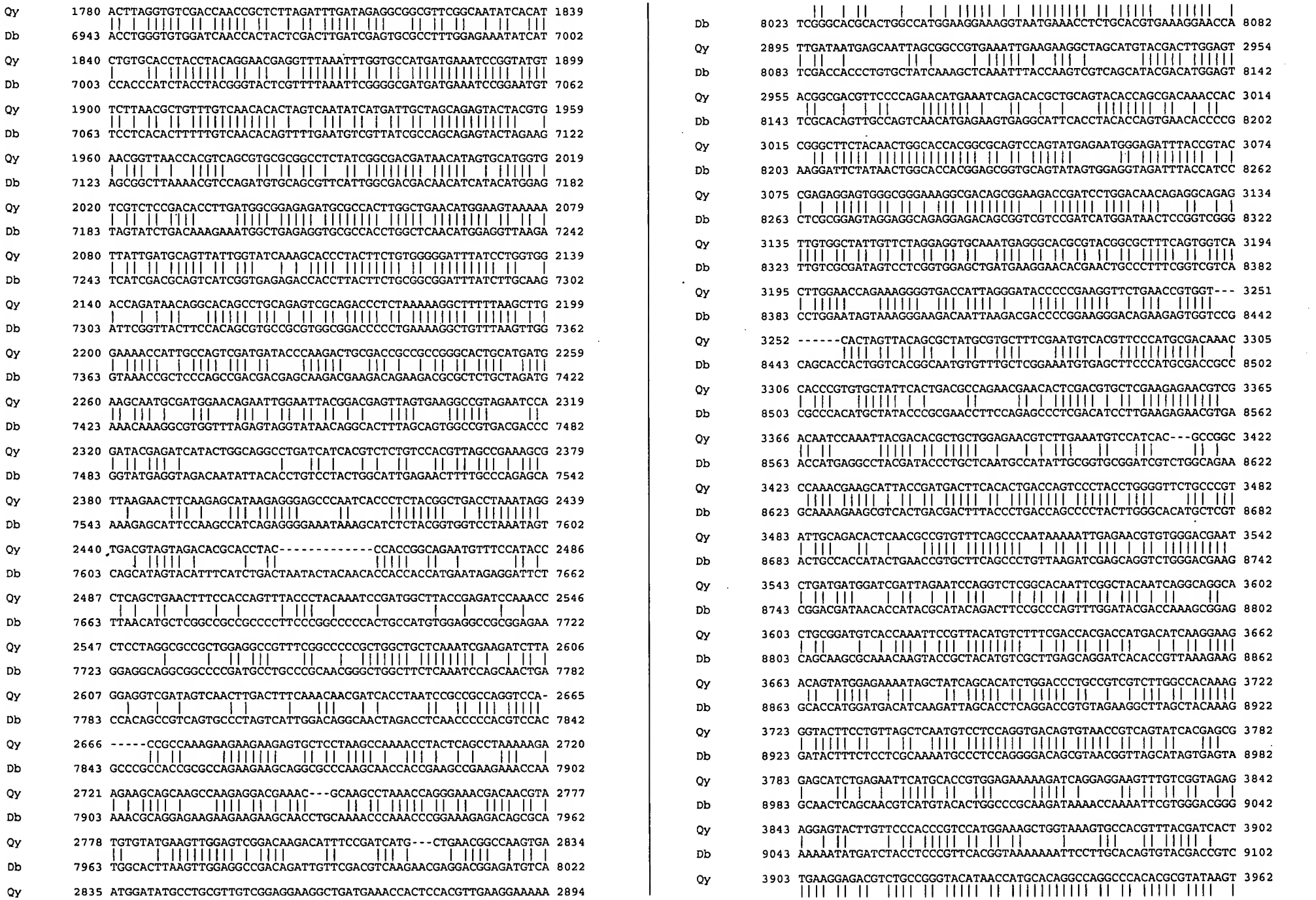
```
Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
Db       5746 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5805

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
Db       5806 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTCCTGG 5865

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
Db       5866 AAAGAATTCATGCCCCGGTGCTCGACACGTCGAAAGAGGAACAACTCAAACTCAGGTACC 5925

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
Db       5926 AGATGATGCCCACCGAAGCCAACAAAAGTAGGTACCAGTCTCGTAAAGTAGAAAATCAGA 5985

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
Db       5986 AAGCCATAACCACTGAGCGACTACTGTCAGGACTACGACTGTATAACTCTGCCACAG--- 6042

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
Db       6043 ATCAGCCAGAATGCTATAAGATCACCTATCCGAAACCATTGTACTCCAGTAGCGTACCGG 6102

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
Db       6103 CGAACTACTCCGATCCACAGTTCGCTGTAGCTGTCTGTAACAACTATCTGCATGAGAACT 6162

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
Db       6163 ATCCGACAGTAGCATCTTATCAGATTACTGACGAGTACGATGCTTACTTGGATATGGTAG 6222

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
Db       6223 ACGGGACAGTCGCCTGCCTGGATACTGCAACCTTCTGCCCCGCTAAGCTTAGAAGTTACC 6282

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
Db       6283 CGAAAAAACATGAGTATAGAGCCCCGAATATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6342

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
Db       6343 ACACGCTACAAAATGTGCTCATTGCCGCAACTAAAAGAAATTGCAACGTCACGCAGATGC 6402

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
Db       6403 GTGAACTGCCAACACTGGACTCAGCGACATTCAATGTCGAATGCTTTCGAAAATATGCAT 6462

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
Db       6463 GTAATGACGAGTATTGGGAGGAGTTCGCTCGGAAGCCAATTAGGATTACCACTGAGTTTG 6522

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
Db       6523 TCACCGCATATGTAGCTAGACTGAAAGGCCCTAAGGCCGCCGCACTATTTGCAAAGACGT 6582

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
Db       6583 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6642

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
Db       6643 ACGTGAAAGTTACACCAGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6702

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
Db       6703 AAGCCGCAGAACCCCTGGCGACTGCTTACTTATGCGGGATTCACCGGGAATTAGTGCGTA 6762

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
Db       6763 GGCTTACGGCCGTCTTGCTTCCAAACATTCACACGCTTTTTGACATGTCGGCGGAGGATT 6822

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
Db       6823 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGCGACCCGGTACTGGAGACGGATATCG 6882

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
```

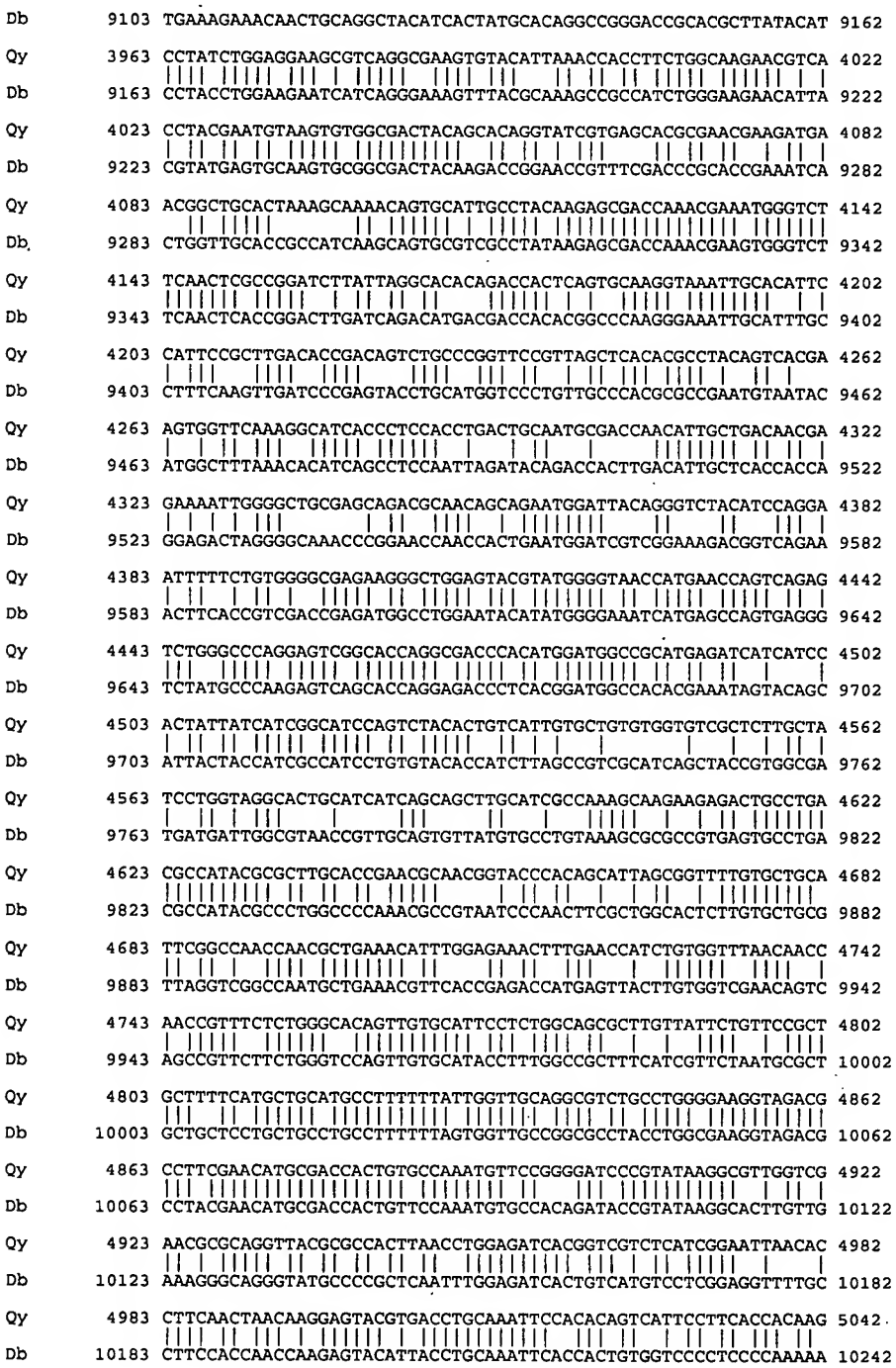

Db 6883 CATCATTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGTCTGATGATCTTGGAGG 6942
Qy 1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
Db 6943 ACCTGGGTGTGGATCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 7002
Qy 1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
Db 7003 CCACCCATCTACCTACGGGTACTCGTTTTAAATTCGGGGCGATGATGAAATCCGGAATGT 7062
Qy 1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
Db 7063 TCCTCACACTTTTTGTCAACACAGTTTTGAATGTCGTTATCGCCAGCAGAGTACTAGAAG 7122
Qy 1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
Db 7123 AGCGGCTTAAAACGTCCAGATGTGCAGCGTTCATTGGCGACGACAACATCATACATGGAG 7182
Qy 2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
Db 7183 TAGTATCTGACAAAGAAATGGCTGAGAGGTGCGCCACCTGGCTCAACATGGAGGTTAAGA 7242
Qy 2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139
Db 7243 TCATCGACGCAGTCATCGGTGAGAGACCACCTTACTTCTGCGGCGGATTTATCTTGCAAG 7302
Qy 2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199
Db 7303 ATTCGGTTACTTCCACAGCGTGCCGCGTGGCGGACCCCCTGAAAAGGCTGTTTAAGTTGG 7362
Qy 2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
Db 7363 GTAAACCGCTCCCAGCCGACGACGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 7422
Qy 2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319
Db 7423 AAACAAAGGCGTGGTTTAGAGTAGGTATAACAGGCACTTTAGCAGTGGCCGTGACGACCC 7482
Qy 2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379
Db 7483 GGTATGAGGTAGACAATATTACACCTGTCCTACTGGCATTGAGAACTTTTGCCCAGAGCA 7542
Qy 2380 TTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGG 2439
Db 7543 AAAGAGCATTCCAAGCCATCAGAGGGGAAATAAAGCATCTCTACGGTGGTCCTAAATAGT 7602
Qy 2440 TGACGTAGTAGACACGCACCTAC----------CCACCGGCAGAATGTTTCCATACC 2486
Db 7603 CAGCATAGTACATTTCATCTGACTAATACTACAACACCACCACCATGAATAGAGGATTCT 7662
Qy 2487 CTCAGCTGAACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACC 2546
Db 7663 TTAACATGCTCGGCCGCCGCCCCTTCCCGGCCCCCACTGCCATGTGGAGGCCGCGGAGAA 7722
Qy 2547 CTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTA 2606
Db 7723 GGAGGCAGGCGGCCCCGATGCCTGCCCGCAACGGGCTGGCTTCTCAAATCCAGCAACTGA 7782
Qy 2607 GGAGGTCGATAGTCAACTTGACTTTCAAACAACGATCACCTAATCCGCCGCCAGGTCCA- 2665
Db 7783 CCACAGCCGTCAGTGCCCTAGTCATTGGACAGGCAACTAGACCTCAACCCCCACGTCCAC 7842
Qy 2666 -----CCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGA 2720
Db 7843 GCCCGCCACCGCGCCAGAAGAAGCAGGCGCCCAAGCAACCACCGAAGCCGAAGAAACCAA 7902
Qy 2721 AGAAGCAGCAAGCCAAGAGGACGAAAC---GCAAGCCTAAACCAGGGAAACGACAACGTA 2777
Db 7903 AAACGCAGGAGAAGAAGAAGAAGCAACCTGCAAAACCCAAACCCGGAAAGAGACAGCGCA 7962
Qy 2778 TGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATCATG---CTGAACGGCCAAGTGA 2834
Db 7963 TGGCACTTAAGTTGGAGGCCGACAGATTGTTCGACGTCAAGAACGAGGACGGAGATGTCA 8022
Qy 2835 ATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACTCCACGTTGAAGGAAAAA 2894
Db 8023 TCGGGCACGCACTGGCCATGGAAGGAAAGGTAATGAAACCTCTGCACGTGAAAGGAACCA 8082
Qy 2895 TTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAGCATGTACGACTTGGAGT 2954
Db 8083 TCGACCACCCTGTGCTATCAAAGCTCAAATTTACCAAGTCGTCAGCATACGACATGGAGT 8142
Qy 2955 ACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTACACCAGCGACAAACCAC 3014
Db 8143 TCGCACAGTTGCCAGTCAACATGAGAAGTGAGGCATTCACCTACACCAGTGAACACCCCG 8202
Qy 3015 CGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAATGGGAGATTTACCGTAC 3074
Db 8203 AAGGATTCTATAACTGGCACCACGGAGCGGTGCAGTATAGTGGAGGTAGATTTACCATCC 8262
Qy 3075 CGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGACAACAGAGGCAGAG 3134
Db 8263 CTCGCGGAGTAGGAGGCAGAGGAGACAGCGGTCGTCCGATCATGGATAACTCCGGTCGGG 8322
Qy 3135 TTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGCGCTTTCAGTGGTCA 3194
Db 8323 TTGTCGCGATAGTCCTCGGTGGAGCTGATGAAGGAACACGAACTGCCCTTTCGGTCGTCA 8382
Qy 3195 CTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTCTGAACCGTGGT--- 3251
Db 8383 CCTGGAATAGTAAAGGGAAGACAATTAAGACGACCCCGGAAGGGACAGAAGAGTGGTCCG 8442
Qy 3252 ------CACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGACAAAC 3305
Db 8443 CAGCACCACTGGTCACGGCAATGTGTTTGCTCGGAAATGTGAGCTTCCCATGCGACCGCC 8502
Qy 3306 CACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAACGTCG 3365
Db 8503 CGCCCACATGCTATACCCGCGAACCTTCCAGAGCCCTCGACATCCTTGAAGAGAACGTGA 8562
Qy 3366 ACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAATGTCCATCAC---GCCGGC 3422
Db 8563 ACCATGAGGCCTACGATACCCTGCTCAATGCCATATTGCGGTGCGGATCGTCTGGCAGAA 8622
Qy 3423 CCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCCCGT 3482
Db 8623 GCAAAAGAAGCGTCACTGACGACTTTACCCTGACCAGCCCCTACTTGGGCACATGCTCGT 8682
Qy 3483 ATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACGAAT 3542
Db 8683 ACTGCCACCATACTGAACCGTGCTTCAGCCCTGTTAAGATCGAGCAGGTCTGGGACGAAG 8742
Qy 3543 CTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAGGCA 3602
Db 8743 CGGACGATAACACCATACGCATACAGACTTCCGCCCAGTTTGGATACGACCAAAGCGGAG 8802
Qy 3603 CTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAAGGAAG 3662
Db 8803 CAGCAAGCGCAAACAAGTACCGCTACATGTCGCTTGAGCAGGATCACACCGTTAAAGAAG 8862
Qy 3663 ACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCACAAAG 3722
Db 8863 GCACCATGGATGACATCAAGATTAGCACCTCAGGACCGTGTAGAAGGCTTAGCTACAAAG 8922
Qy 3723 GGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCACGAGCG 3782
Db 8923 GATACTTTCTCCTCGCAAAATGCCCTCCAGGGGACAGCGTAACGGTTAGCATAGTGAGTA 8982
Qy 3783 GAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTAGAG 3842
Db 8983 GCAACTCAGCAACGTCATGTACACTGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGG 9042
Qy 3843 AGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATCACT 3902
Db 9043 AAAAATATGATCTACCTCCCGTTCACGGTAAAAAAATTCCTTGCACAGTGTACGACCGTC 9102


Qy 3903 TGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAAGT 3962
Db 9103 TGAAAGAAACAACTGCAGGCTACATCACTATGCACAGGCCGGGACCGCACGCTTATACAT 9162
Qy 3963 CCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGTCA 4022
Db 9163 CCTACCTGGAAGAATCATCAGGGAAAGTTTACGCAAAGCCGCCATCTGGGAAGAACATTA 9222
Qy 4023 CCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGATGA 4082
Db 9223 CGTATGAGTGCAAGTGCGGCGACTACAAGACCGGAACCGTTTCGACCCGCACCGAAATCA 9282
Qy 4083 ACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGTCT 4142
Db 9283 CTGGTTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAACGAAGTGGGTCT 9342
Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9343 TCAACTCACCGGACTTGATCAGACATGACGACCACACGGCCCAAGGGAAATTGCATTTGC 9402
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9403 CTTTCAAGTTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAATGTAATAC 9462
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9463 ATGGCTTTAAACACATCAGCCTCCAATTAGATACAGACCACTTGACATTGCTCACCACCA 9522
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9523 GGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCGTCGGAAAGACGGTCAGAA 9582
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9583 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGAAATCATGAGCCAGTGAGGG 9642
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db 9643 TCTATGCCCAAGAGTCAGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9702
Qy 4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db 9703 ATTACTACCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTACCGTGGCGA 9762
Qy 4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db 9763 TGATGATTGGCGTAACCGTTGCAGTGTTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9822
Qy 4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db 9823 CGCCATACGCCCTGGCCCCAAACGCCGTAATCCCAACTTCGCTGGCACTCTTGTGCTGCG 9882
Qy 4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db 9883 TTAGGTCGGCCAATGCTGAAACGTTCACCGAGACCATGAGTTACTTGTGGTCGAACAGTC 9942
Qy 4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
Db 9943 AGCCGTTCTTCTGGGTCCAGTTGTGCATACCTTTGGCCGCTTTCATCGTTCTAATGCGCT 10002
Qy 4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
Db 10003 GCTGCTCCTGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10062
Qy 4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
Db 10063 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10122
Qy 4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db 10123 AAAGGGCAGGGTATGCCCCGCTCAATTTGGAGATCACTGTCATGTCCTCGGAGGTTTTGC 10182

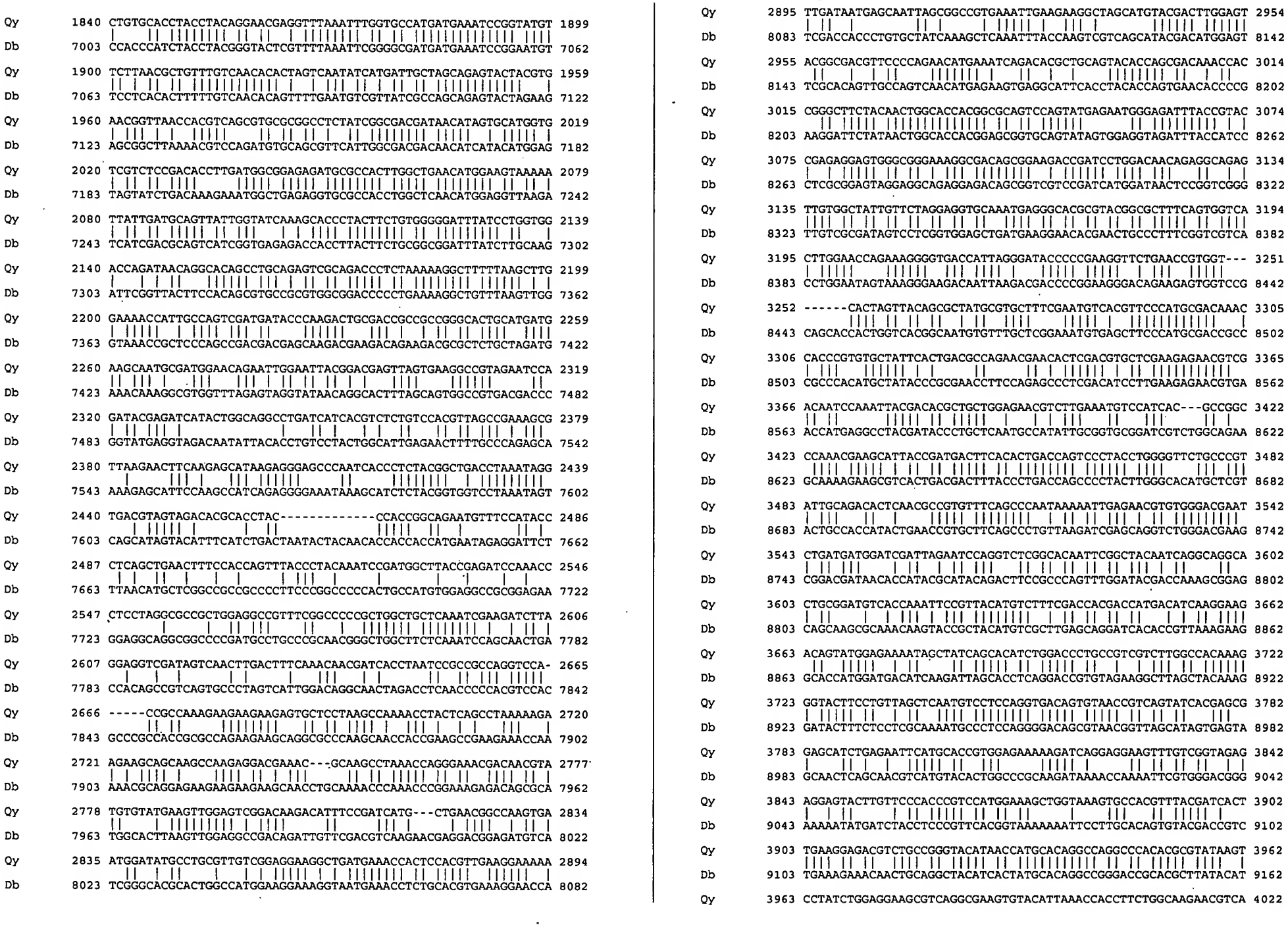

Qy 4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db 10183 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCAAAAA 10242
Qy 5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db 10243 TCAAATGCTGCGGCTCCTTGGAATGTCAGCCGGCCGCTCATGCAGACTATACCTGCAAGG 10302
Qy 5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db 10303 TCTTCGGAGGGGTCTACCCCTTTATGTGGGGAGGAGCGCAATGTTTTTGCGACAGTGAGA 10362
Qy 5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db 10363 ACAGCCAGATGAGTGAGGCGTACGTCGAACTGTCAGCAGATTGCGCGTCTGACCACGCGC 10422
Qy 5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db 10423 AGGCGATTAAGGTGCACACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10482
Qy 5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db 10483 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACTTGA 10542
Qy 5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db 10543 AAGTCATAGCTGGACCAATTTCAGCATCGTTTACGCCATTCGATCATAAGGTCGTTATCC 10602
Qy 5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db 10603 ATCGCGGCCTGGTGTACAACTATGACTTCCCGGAATATGGAGCGATGAAACCAGGAGCGT 10662
Qy 5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db 10663 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAGGATCTCATCGCCAGCACAGACATTA 10722
Qy 5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db 10723 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCAGGAT 10782
Qy 5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db 10783 TTGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCACCTTTCGGGTGTA 10842
Qy 5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db 10843 AGATTGCAGTAAATCCGCTCCGAGCGGTGGACTGTTCATACGGGAACATTCCCATTTCTA 10902
Qy 5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db 10903 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10962
Qy 5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db 10963 AATGTGAAGTCAGTGAGTGCACTTATTCAGCAGACTTCGGCGGGATGGCCACCCTGCAGT 11022
Qy 5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db 11023 ATGTATCCGACCGCGAAGGTCAATGCCCCGTACATTCGCATTCGAGCACAGCAACTCTCC 11082
Qy 5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db 11083 AAGAGTCGACAGTACATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTTAGCACCGCGA 11142
Qy 5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db 11143 GTCCACAGGCGAACTTTATCGTATCGCTGTGTGGGAAGAAGACAACATGCAATGCAGAAT 11202
Qy 6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db 11203 GTAAACCACCAGCTGACCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTTCAAG 11262
Qy 6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122

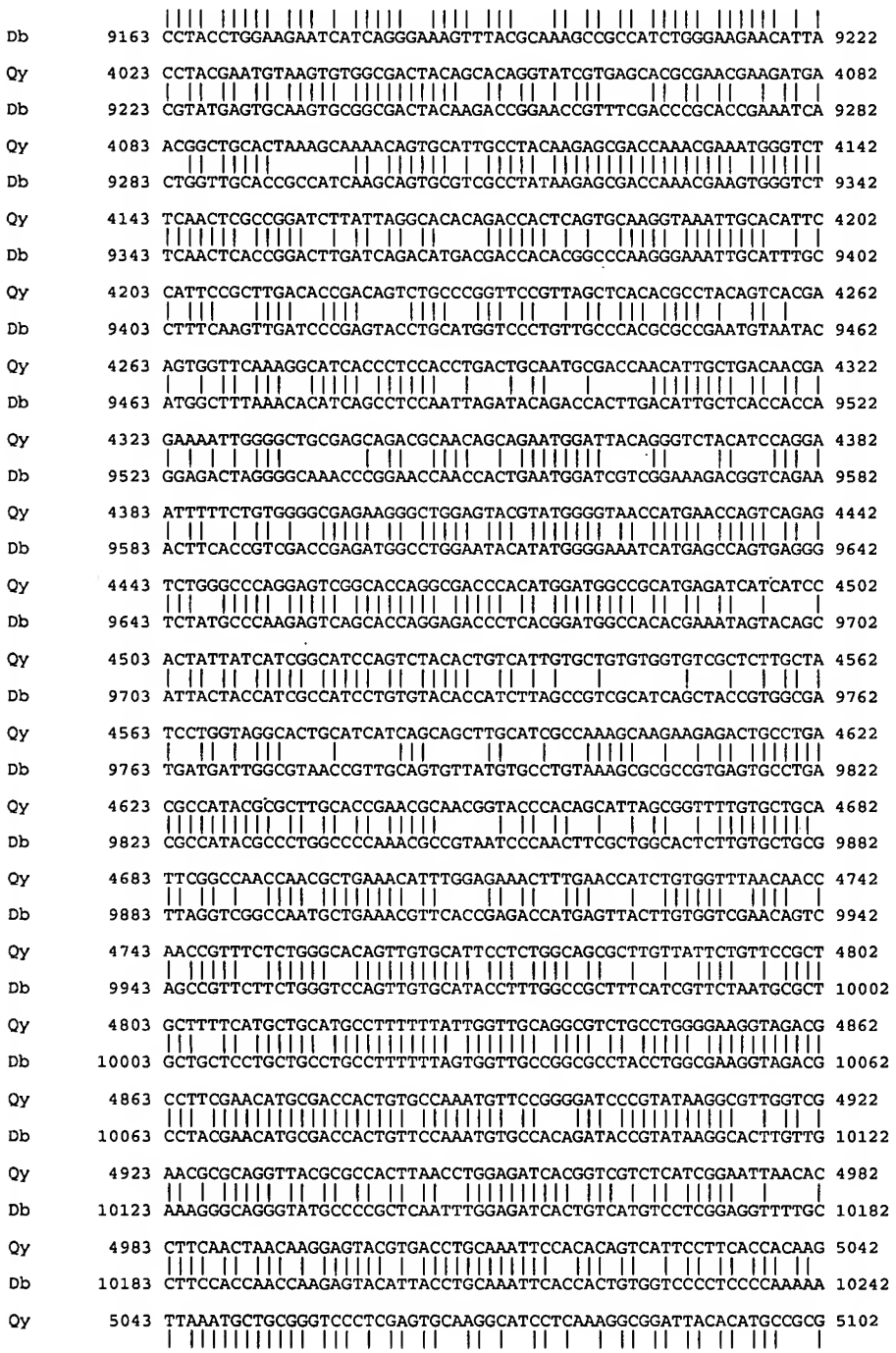
```
Db      11263 CCGCCATCTCAAAAACATCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11322

Qy       6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182

Db      11323 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11382

Qy       6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241

Db      11383 GACCGCTACGCCCCAATGATCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11442

Qy       6242 GCATAATGCCACGCGCCGCT 6261

Db      11443 GCATAATGCATCAGGCTGGT 11462


RESULT 10
US-09-102-248-8
; Sequence 8, Application US/09102248
; Patent No. 6008035
;  GENERAL INFORMATION:
;    APPLICANT:  Johnston, Robert E.
;    APPLICANT:  Davis, Nancy L.
;    APPLICANT:  Simpson, Dennis A.
;    TITLE OF INVENTION:  System for the In Vivo Delivery and
;    TITLE OF INVENTION:  Expression of Heterologous Genes in the Bone Marrow
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson, P.A.
;      STREET:  1211 East Morehead Street
;      CITY:  Charlotte
;      STATE:  No. 6008035th Carolina
;      COUNTRY:  USA
;      ZIP:  28234
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/102,248
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/801,263
;      FILING DATE:  19-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sibley, Kenneth D.
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5470-147
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-420-2200
;      TELEFAX:  919-881-3175
;   INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11703 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-102-248-8

  Query Match             36.8%;  Score 2388.4;  DB 3;  Length 11703;
  Best Local Similarity   64.7%;  Pred. No. 0;
  Matches 3698;  Conservative    0;  Mismatches 1981;  Indels   41;  Gaps    8;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639

Db       5746 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5805

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699

Db       5806 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTCCTGG 5865

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759

Db       5866 AAAGAATTCATGCCCCGGTGCTCGACACGTCGAAAGAGGAACAACTCAAACTCAGGTACC 5925

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819

Db       5926 AGATGATGCCCACCGAAGCCAACAAAAGTAGGTACCAGTCTCGTAAAGTAGAAAATCAGA 5985

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879

Db       5986 AAGCCATAACCACTGAGCGACTACTGTCAGGACTACGACTGTATAACTCTGCCACAG--- 6042

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939

Db       6043 ATCAGCCAGAATGCTATAAGATCACCTATCCGAAACCATTGTACTCCAGTAGCGTACCGG 6102

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999

Db       6103 CGAACTACTCCGATCCACAGTTCGCTGTAGCTGTCTGTAACAACTATCTGCATGAGAACT 6162

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059

Db       6163 ATCCGACAGTAGCATCTTATCAGATTACTGACGAGTACGATGCTTACTTGGATATGGTAG 6222

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119

Db       6223 ACGGGACAGTCGCCTGCCTGGATACTGCAACCTTCTGCCCCGCTAAGCTTAGAAGTTACC 6282

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179

Db       6283 CGAAAAAACATGAGTATAGAGCCCCGAATATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6342

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239

Db       6343 ACACGCTACAAAATGTGCTCATTGCCGCAACTAAAAGAAATTGCAACGTCACGCAGATGC 6402

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299

Db       6403 GTGAACTGCCAACACTGGACTCAGCGACATTCAATGTCGAATGCTTTCGAAAATATGCAT 6462

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359

Db       6463 GTAATGACGAGTATTGGGAGGAGTTCGCTCGGAAGCCAATTAGGATTACCACTGAGTTTG 6522

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419

Db       6523 TCACCGCATATGTAGCTAGACTGAAAGGCCCTAAGGCCGCCGCACTATTTGCAAAGACGT 6582

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479

Db       6583 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6642

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539

Db       6643 ACGTGAAAGTTACACCAGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6702

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599

Db       6703 AAGCCGCAGAACCCCTGGCGACTGCTTACTTATGCGGGATTCACCGGGAATTAGTGCGTA 6762

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659

Db       6763 GGCTTACGGCCGTCTTGCTTCCAAACATTCACACGCTTTTTGACATGTCGGCGGAGGATT 6822

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719

Db       6823 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGCGACCCGGTACTGGAGACGGATATCG 6882

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779

Db       6883 CATCATTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGTCTGATGATCTTGGAGG 6942
```

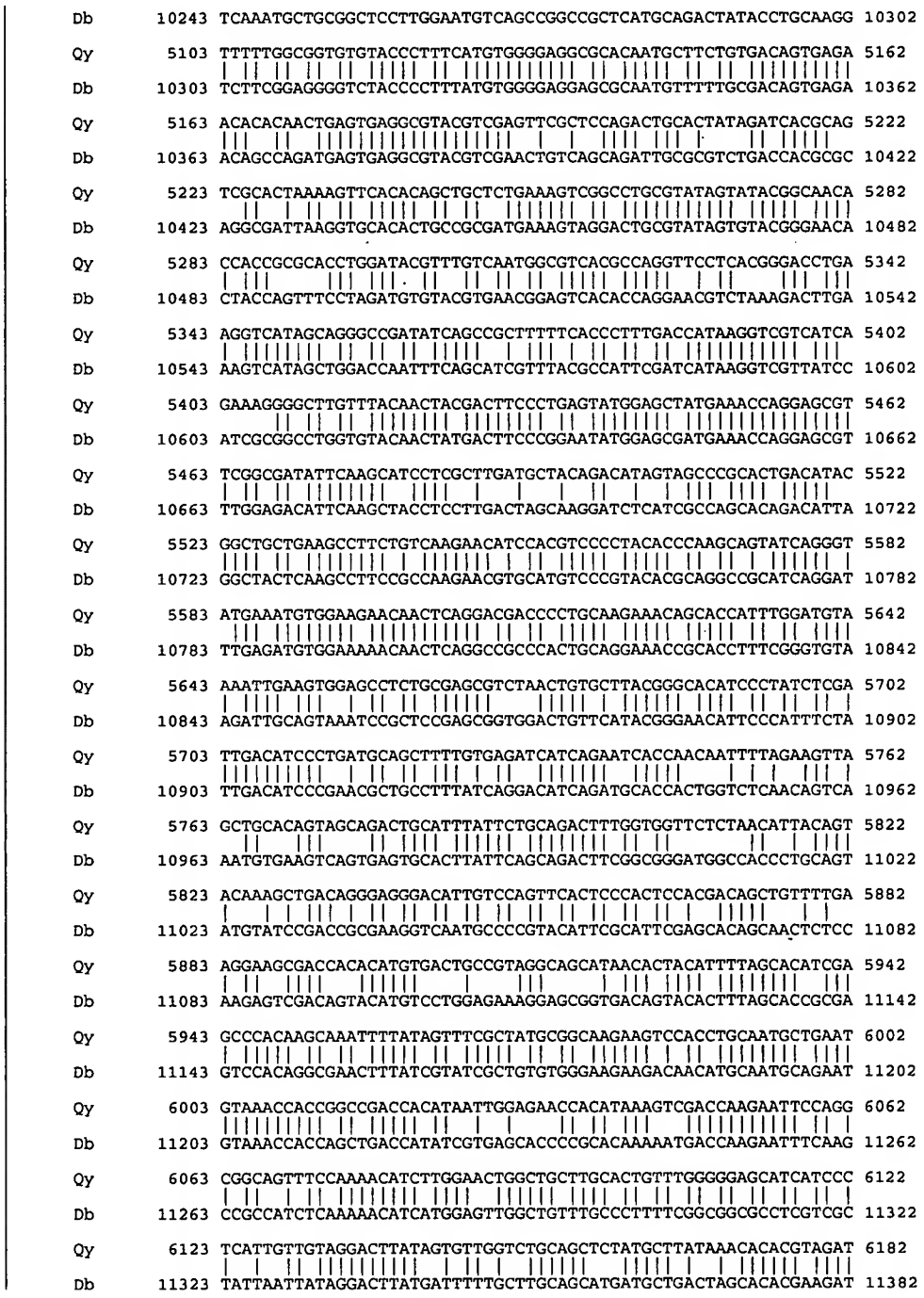



Db 9103 TGAAAGAAACAACTGCAGGCTACATCACTATGCACAGGCCGGGACCGCACGCTTATACAT 9162
Qy 3963 CCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGTCA 4022
Db 9163 CCTACCTGGAAGAATCATCAGGGAAAGTTTACGCAAAGCCGCCATCTGGGAAGAACATTA 9222
Qy 4023 CCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGATGA 4082
Db 9223 CGTATGAGTGCAAGTGCGGCGACTACAAGACCGGAACCGTTTCGACCCGCACCGAAATCA 9282
Qy 4083 ACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGTCT 4142
Db 9283 CTGGTTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAACGAAGTGGGTCT 9342
Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9343 TCAACTCACCGGACTTGATCAGACATGACGACCACACGGCCCAAGGGAAATTGCATTTGC 9402
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9403 CTTTCAAGTTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAATGTAATAC 9462
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9463 ATGGCTTTAAACACATCAGCCTCCAATTAGATACAGACCACTTGACATTGCTCACCACCA 9522
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9523 GGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCGTCGGAAAGACGGTCAGAA 9582
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9583 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGAAATCATGAGCCAGTGAGGG 9642
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db 9643 TCTATGCCCAAGAGTCAGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9702
Qy 4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db 9703 ATTACTACCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTACCGTGGCGA 9762
Qy 4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db 9763 TGATGATTGGCGTAACCGTTGCAGTGTTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9822
Qy 4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db 9823 CGCCATACGCCCTGGCCCCAAACGCCGTAATCCCAACTTCGCTGGCACTCTTGTGCTGCG 9882
Qy 4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db 9883 TTAGGTCGGCCAATGCTGAAACGTTCACCGAGACCATGAGTTACTTGTGGTCGAACAGTC 9942
Qy 4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
Db 9943 AGCCGTTCTTCTGGGTCCAGTTGTGCATACCTTTGGCCGCTTTCATCGTTCTAATGCGCT 10002
Qy 4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
Db 10003 GCTGCTCCTGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10062
Qy 4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
Db 10063 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10122
Qy 4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db 10123 AAAGGGCAGGGTATGCCCCGCTCAATTTGGAGATCACTGTCATGTCCTCGGAGGTTTTGC 10182
Qy 4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db 10183 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCAAAAA 10242
Qy 5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db 10243 TCAAATGCTGCGGCTCCTTGGAATGTCAGCCGGCCGCTCATGCAGACTATACCTGCAAGG 10302
Qy 5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db 10303 TCTTCGGAGGGGTCTACCCCTTTATGTGGGGAGGAGCGCAATGTTTTTGCGACAGTGAGA 10362
Qy 5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db 10363 ACAGCCAGATGAGTGAGGCGTACGTCGAACTGTCAGCAGATTGCGCGTCTGACCACGCGC 10422
Qy 5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db 10423 AGGCGATTAAGGTGCACACTGCCGCGATGAAAGTAGGACTGCGTATAGTGTACGGGAACA 10482
Qy 5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db 10483 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACTTGA 10542
Qy 5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db 10543 AAGTCATAGCTGGACCAATTTCAGCATCGTTTACGCCATTCGATCATAAGGTCGTTATCC 10602
Qy 5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db 10603 ATCGCGGCCTGGTGTACAACTATGACTTCCCGGAATATGGAGCGATGAAACCAGGAGCGT 10662
Qy 5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db 10663 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAGGATCTCATCGCCAGCACAGACATTA 10722
Qy 5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db 10723 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCAGGAT 10782
Qy 5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db 10783 TTGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCACCTTTCGGGTGTA 10842
Qy 5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db 10843 AGATTGCAGTAAATCCGCTCCGAGCGGTGGACTGTTCATACGGGAACATTCCCATTTCTA 10902
Qy 5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db 10903 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10962
Qy 5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db 10963 AATGTGAAGTCAGTGAGTGCACTTATTCAGCAGACTTCGGCGGGATGGCCACCCTGCAGT 11022
Qy 5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db 11023 ATGTATCCGACCGCGAAGGTCAATGCCCCGTACATTCGCATTCGAGCACAGCAACTCTCC 11082
Qy 5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db 11083 AAGAGTCGACAGTACATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTTAGCACCGCGA 11142
Qy 5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db 11143 GTCCACAGGCGAACTTTATCGTATCGCTGTGTGGGAAGAAGACAACATGCAATGCAGAAT 11202
Qy 6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db 11203 GTAAACCACCAGCTGACCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTTCAAG 11262
Qy 6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db 11263 CCGCCATCTCAAAAACATCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11322

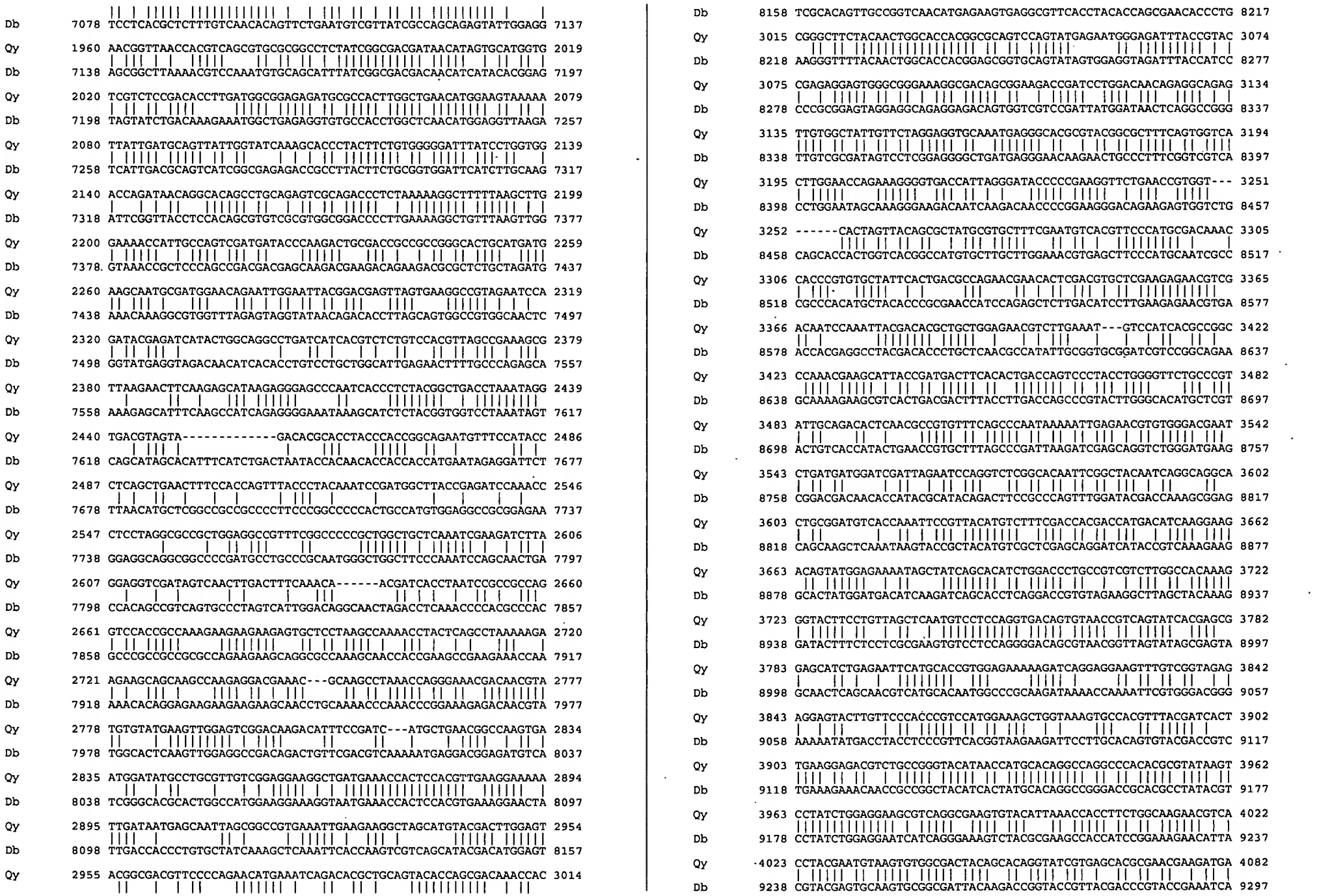
```
Qy       6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
Db      11323 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11382

Qy       6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
Db      11383 GACCGCTACGCCCCAATGATCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11442

Qy       6242 GCATAATGCCACGCGCCGCT 6261
Db      11443 GCATAATGCATCAGGCTGGT 11462


RESULT 11
US-09-367-764-8
; Sequence 8, Application US/09367764
; Patent No. 6583121
;  GENERAL INFORMATION:
;    APPLICANT:  Johnston, Robert E.
;    APPLICANT:  Davis, Nancy L.
;    APPLICANT:  Simpson, Dennis A.
;    TITLE OF INVENTION:  System for the In Vivo Delivery and
;    TITLE OF INVENTION:  Expression of Heterologous Genes in the Bone Marrow
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson, P.A.
;      STREET:  1211 East Morehead Street
;      CITY:  Charlotte
;      STATE:  No. 6583121th Carolina
;      COUNTRY:  USA
;      ZIP:  28234
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/367,764
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/801,263
;      FILING DATE:  19-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sibley, Kenneth D.
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5470-147
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-420-2200
;      TELEFAX:  919-881-3175
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11703 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-367-764-8

  Query Match             36.8%;  Score 2388.4;  DB 4;  Length 11703;
  Best Local Similarity   64.7%;  Pred. No. 0;
  Matches 3698;  Conservative    0;  Mismatches 1981;  Indels   41;  Gaps    8;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
Db       5746 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5805

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
Db       5806 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTCCTGG 5865

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759

Db       5866 AAAGAATTCATGCCCCGGTGCTCGACACGTCGAAAGAGGAACAACTCAAACTCAGGTACC 5925

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
Db       5926 AGATGATGCCCACCGAAGCCAACAAAAGTAGGTACCAGTCTCGTAAAGTAGAAAATCAGA 5985

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
Db       5986 AAGCCATAACCACTGAGCGACTACTGTCAGGACTACGACTGTATAACTCTGCCACAG--- 6042

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
Db       6043 ATCAGCCAGAATGCTATAAGATCACCTATCCGAAACCATTGTACTCCAGTAGCGTACCGG 6102

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
Db       6103 CGAACTACTCCGATCCACAGTTCGCTGTAGCTGTCTGTAACAACTATCTGCATGAGAACT 6162

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
Db       6163 ATCCGACAGTAGCATCTTATCAGATTACTGACGAGTACGATGCTTACTTGGATATGGTAG 6222

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
Db       6223 ACGGGACAGTCGCCTGCCTGGATACTGCAACCTTCTGCCCCGCTAAGCTTAGAAGTTACC 6282

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
Db       6283 CGAAAAAACATGAGTATAGAGCCCCGAATATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6342

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
Db       6343 ACACGCTACAAAATGTGCTCATTGCCGCAACTAAAAGAAATTGCAACGTCACGCAGATGC 6402

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
Db       6403 GTGAACTGCCAACACTGGACTCAGCGACATTCAATGTCGAATGCTTTCGAAAATATGCAT 6462

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
Db       6463 GTAATGACGAGTATTGGGAGGAGTTCGCTCGGAAGCCAATTAGGATTACCACTGAGTTTG 6522

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
Db       6523 TCACCGCATATGTAGCTAGACTGAAAGGCCCTAAGGCCGCCGCACTATTTGCAAAGACGT 6582

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
Db       6583 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6642

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
Db       6643 ACGTGAAAGTTACACCAGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6702

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
Db       6703 AAGCCGCAGAACCCCTGGCGACTGCTTACTTATGCGGGATTCACCGGGAATTAGTGCGTA 6762

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
Db       6763 GGCTTACGGCCGTCTTGCTTCCAAACATTCACACGCTTTTTGACATGTCGGCGGAGGATT 6822

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
Db       6823 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGCGACCCGGTACTGGAGACGGATATCG 6882

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
Db       6883 CATCATTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGTCTGATGATCTTGGAGG 6942

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
Db       6943 ACCTGGGTGTGGATCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 7002
```

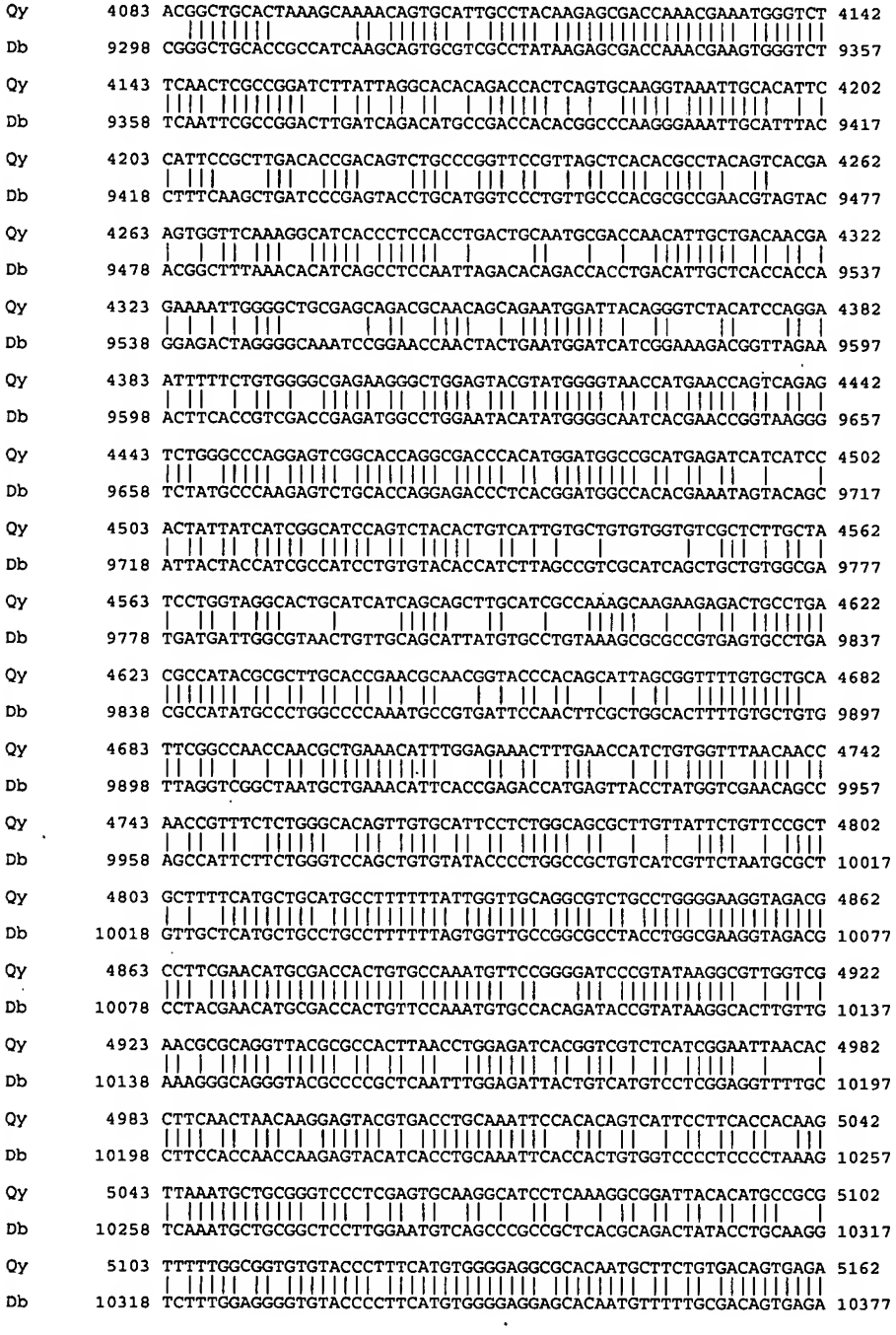




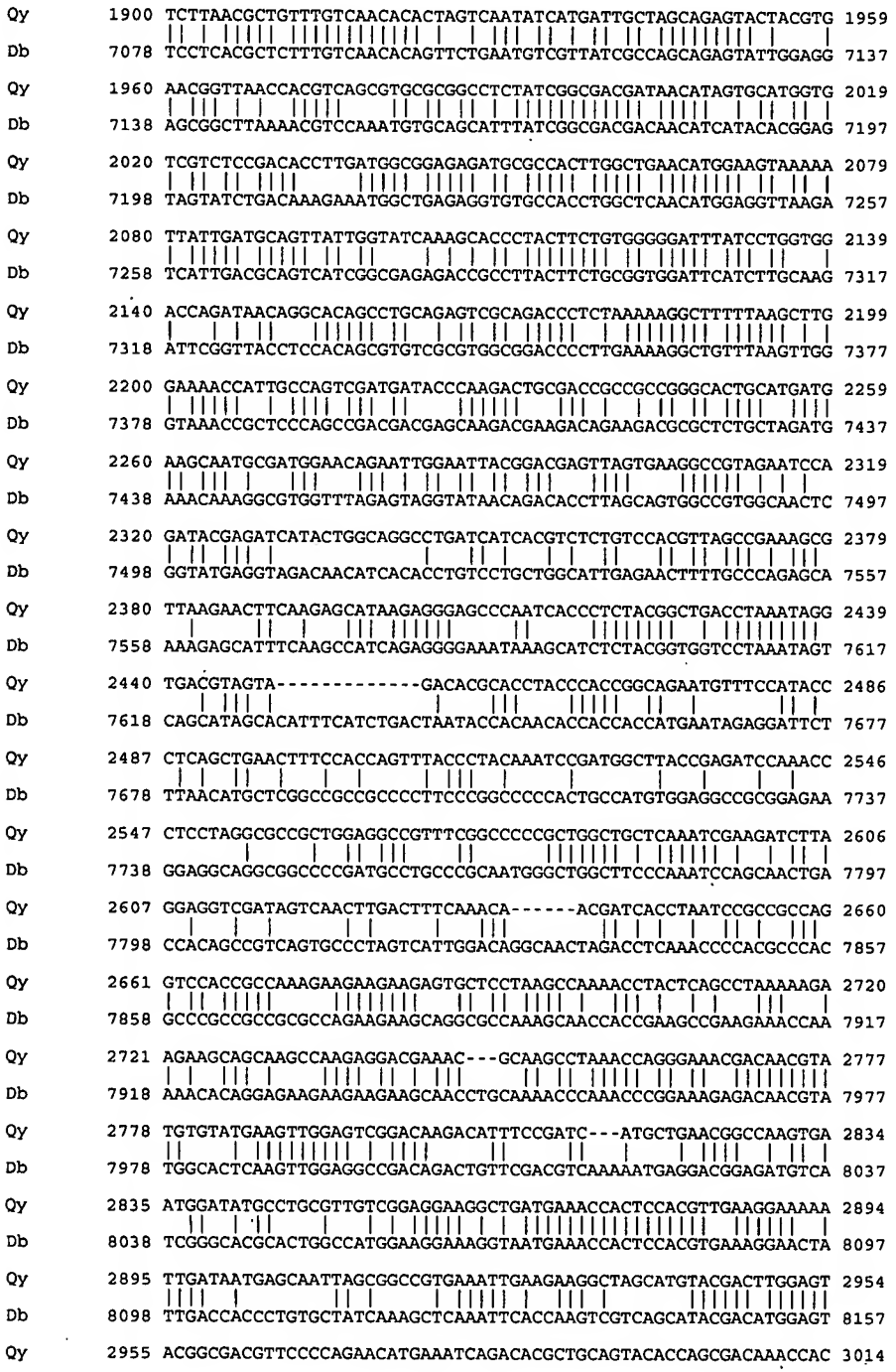


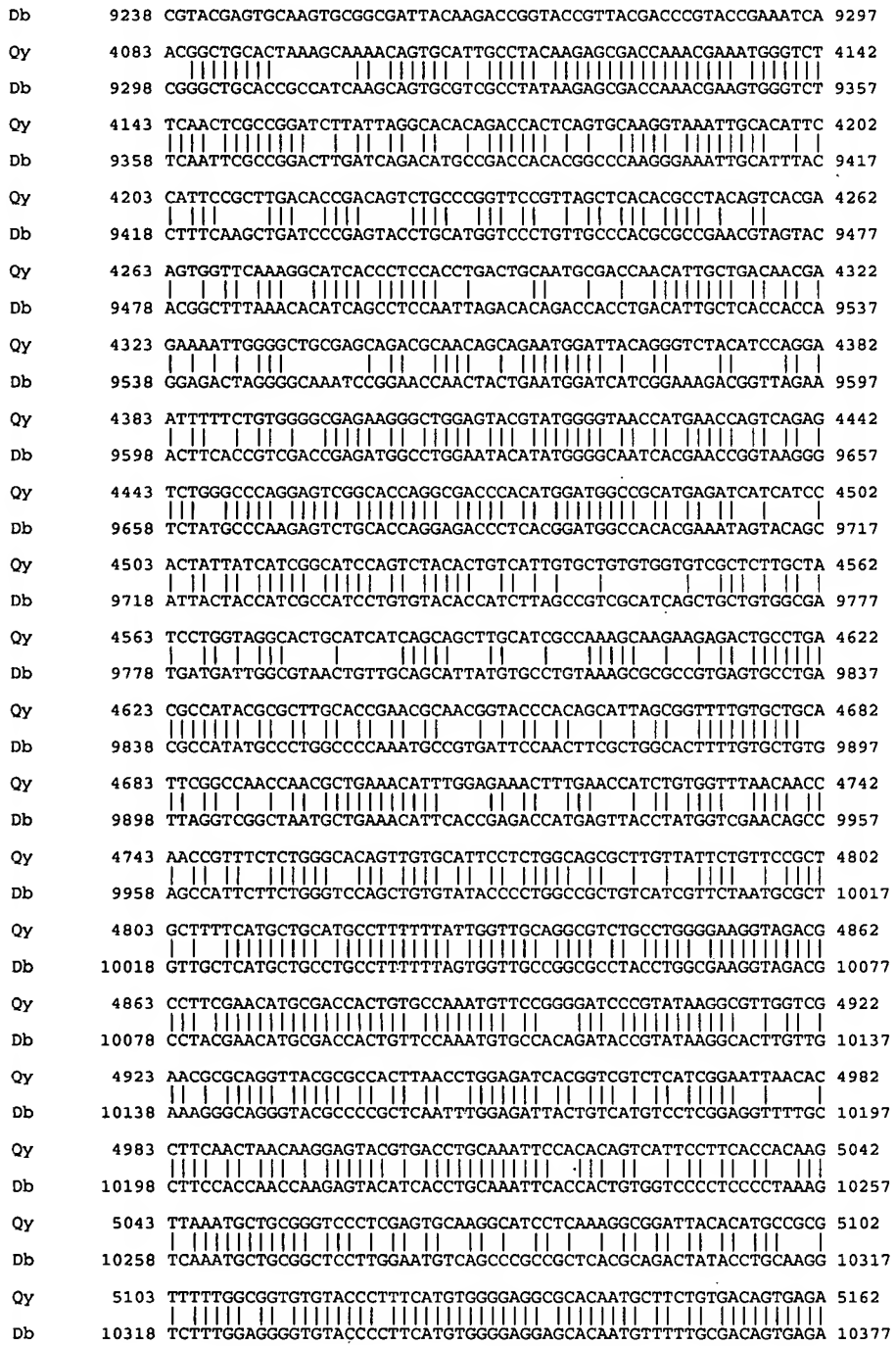
```
Qy       6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
Db      11383 GACCGCTACGCCCCAATGATCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11442

Qy       6242 GCATAATGCCACGCGCCGCT 6261
Db      11443 GCATAATGCATCAGGCTGGT 11462


RESULT 12
US-08-801-263A-4
; Sequence 4, Application US/08801263A
; Patent No. 5811407
;  GENERAL INFORMATION:
;    APPLICANT:  Johnston, Robert E.
;    APPLICANT:  Davis, Nancy L.
;    APPLICANT:  Simpson, Dennis A.
;    TITLE OF INVENTION:  System for the In Vivo Delivery and
;    TITLE OF INVENTION:  Expression of Heterologous Genes in the Bone Marrow
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson, P.A.
;      STREET:  1211 East Morehead Street
;      CITY:  Charlotte
;      STATE:  No. 5811407th Carolina
;      COUNTRY:  USA
;      ZIP:  28234
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/801,263A
;      FILING DATE:  19-FEB-1997
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sibley, Kenneth D.
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5470-147
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-420-2200
;      TELEFAX:  919-881-3175
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11717 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-801-263A-4

  Query Match             36.8%;  Score 2386.4;  DB 1;  Length 11717;
  Best Local Similarity   64.5%;  Pred. No. 0;
  Matches 3745;  Conservative    0;  Mismatches 1991;  Indels   74;  Gaps    9;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
Db       5761 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5820

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
Db       5821 AAATGGAGTCCGTTCTGCAGAATCAGCTTACAGAACCGACCTTGGAGCGCAATGTTCTGG 5880

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
Db       5881 AAAGAATCTACGCCCCGGTGCTCGACACGTCGAAAGAGGAACAGCTCAAACTCAGGTACC 5940

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
Db       5941 AGATGATGCCCACCGAAGCCAACAAAAGCAGGTACCAGTCTAGAAAAGTAGAAAATCAGA 6000

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
Db       6001 AAGCCATAACCACTGAGCGACTGCTTTCAGGGCTACGACTGTATAACTCTGCCACAG--- 6057

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
Db       6058 ATCAGCCAGAATGCTATAAGATCACCTACCCGAAACCATCGTATTCCAGCAGTGTACCGG 6117

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
Db       6118 CGAACTACTCTGACCCAAAGTTTGCTGTAGCTGTTTGCAACAACTATCTGCATGAGAATT 6177

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
Db       6178 ACCCGACGGTAGCATCTTATCAGATCACCGACGAGTACGATGCTTACTTGGATATGGTAG 6237

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
Db       6238 ACGGGACAGTCGCTTGCCTAGATACTGCAACTTTTTGCCCCGCCAAGCTTAGAAGTTACC 6297

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
Db       6298 CGAAAAGACACGAGTATAGAGCCCCAAACACTCGCAGTGCGGTTCCATCAGCGATGCAGA 6357

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
Db       6358 ACACGTTGCAAAACGTGCTCATTGCCGCGACTAAAAGAAACTGCAACGTCACACAAATGC 6417

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
Db       6418 GTGAATTGCCAACACTGGACTCAGCGACATTCAACGTTGAATGCTTTCGAAAATATGCAT 6477

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
Db       6478 GTAATGACGAGTATTGGGAGGAGTTTGCCCGAAAGCCAATTAGGATCACTACTGAGTTCG 6537

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
Db       6538 TTACCGCATACGTGGCCAGACTGAAAGGCCCTAAGGCCGCCGCACTGTTCGCAAAGACGC 6597

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
Db       6598 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGGTTCGTCATGGACATGAAAAGAG 6657

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
Db       6658 ACGTGAAAGTTACACCTGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGCTAC 6717

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
Db       6718 AAGCCGCAGAACCCCTGGCGACCGCTTACCTGTGCGGGATCCACCGGGAGTTAGTGCGCA 6777

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
Db       6778 GGCTTACAGCCGTCTTGCTACCCAACATTCACACGCTTTTTGACATGTCGGCGGAGGACT 6837

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
Db       6838 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGTGACCCGGTACTGGAGACGGATATCG 6897

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
Db       6898 CCTCGTTCGACAAAAGCCAAGACGACGCTATGGCGTTAACTGGCCTGATGATCTTGGAAG 6957

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
Db       6958 ACCTGGGTGTGGACCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 7017

Qy       1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
Db       7018 CCACCCATCTGCCCACGGGTACCCGTTTCAAATTCGGGGCGATGATGAAATCCGGAATGT 7077

Qy       1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
```



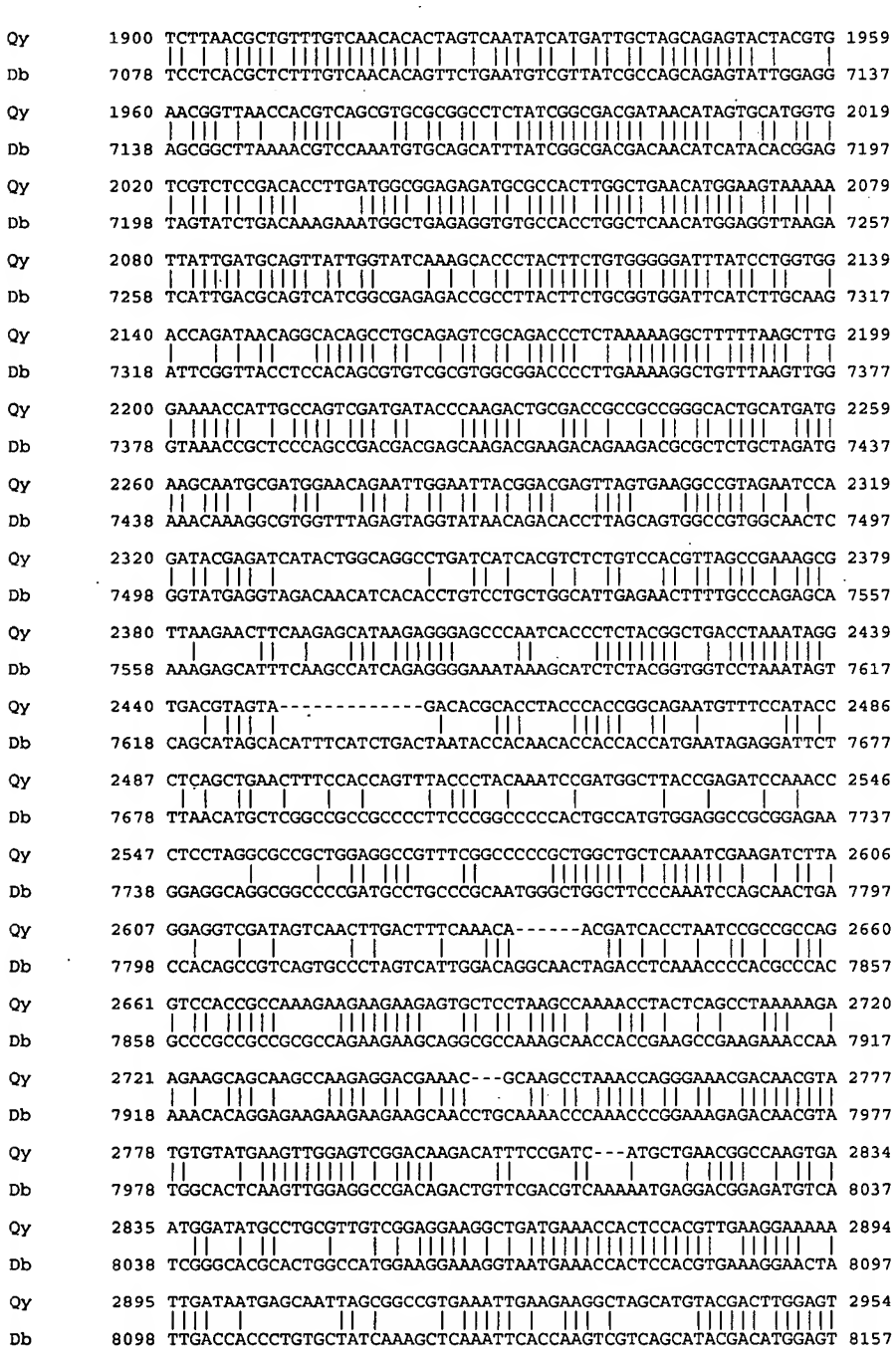

Qy 4083 ACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGTCT 4142
Db 9298 CGGGCTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAACGAAGTGGGTCT 9357
Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9358 TCAATTCGCCGGACTTGATCAGACATGCCGACCACACGGCCCAAGGGAAATTGCATTTAC 9417
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9418 CTTTCAAGCTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAACGTAGTAC 9477
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9478 ACGGCTTTAAACACATCAGCCTCCAATTAGACACAGACCACCTGACATTGCTCACCACCA 9537
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9538 GGAGACTAGGGGCAAATCCGGAACCAACTACTGAATGGATCATCGGAAAGACGGTTAGAA 9597
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9598 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGCAATCACGAACCGGTAAGGG 9657
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db 9658 TCTATGCCCAAGAGTCTGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9717
Qy 4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db 9718 ATTACTACCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTGCTGTGGCGA 9777
Qy 4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db 9778 TGATGATTGGCGTAACTGTTGCAGCATTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9837
Qy 4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db 9838 CGCCATATGCCCTGGCCCCAAATGCCGTGATTCCAACTTCGCTGGCACTTTTGTGCTGTG 9897
Qy 4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db 9898 TTAGGTCGGCTAATGCTGAAACATTCACCGAGACCATGAGTTACCTATGGTCGAACAGCC 9957
Qy 4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
Db 9958 AGCCATTCTTCTGGGTCCAGCTGTGTATACCCCTGGCCGCTGTCATCGTTCTAATGCGCT 10017
Qy 4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
Db 10018 GTTGCTCATGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10077
Qy 4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
Db 10078 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10137
Qy 4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db 10138 AAAGGGCAGGGTACGCCCCGCTCAATTTGGAGATTACTGTCATGTCCTCGGAGGTTTTGC 10197
Qy 4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db 10198 CTTCCACCAACCAAGAGTACATCACCTGCAAATTCACCACTGTGGTCCCCTCCCCTAAAG 10257
Qy 5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db 10258 TCAAATGCTGCGGCTCCTTGGAATGTCAGCCCGCCGCTCACGCAGACTATACCTGCAAGG 10317
Qy 5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db 10318 TCTTTGGAGGGGTGTACCCCTTCATGTGGGGAGGAGCACAATGTTTTTGCGACAGTGAGA 10377
Qy 5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db 10378 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGCAGATTGCGCGACTGACCACGCGC 10437
Qy 5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db 10438 AGGCGATTAAGGTGCATACTGCCGCGATGAAAGTAGGACTACGTATAGTGTACGGGAACA 10497
Qy 5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db 10498 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACCTGA 10557
Qy 5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db 10558 AAGTCATAGCTGGACCAATTTCAGCATCGTTTACACCATTCGATCACAAGGTCGTTATCC 10617
Qy 5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db 10618 ATCGCGGCCTGGTGTACAACTATGACTTCCCGGAATACGGAGCGATGAAACCAGGAGCGT 10677
Qy 5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db 10678 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAAGATCTCATCGCCAGCACAGACATTA 10737
Qy 5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db 10738 GACTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCTGGAT 10797
Qy 5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db 10798 TCGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCCCCTTTCGGGTGCA 10857
Qy 5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db 10858 AGATTGCAGTCAATCCGCTTCGAGCGGTGGACTGCTCATACGGGAACATTCCCATCTCTA 10917
Qy 5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db 10918 TCGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10977
Qy 5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db 10978 AATGTGATGTCAGTGAGTGCACTTACTCAGCGGACTTCGGCGGGATGGCTACCCTGCAGT 11037
Qy 5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db 11038 ATGTATCCGACCGCGAAGGACAATGCCCTGTACATTCGCATTCGAGCACAGCAACCCTCC 11097
Qy 5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db 11098 AAGAGTCGACAGTTCATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTCAGCACCGCGA 11157
Qy 5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db 11158 GCCCACAGGCGAACTTTATTGTATCGCTGTGTGGTAAGAAGACAACATGCAATGCAGAAT 11217
Qy 6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db 11218 GCAAACCACCAGCTGACCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTCCAAG 11277
Qy 6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db 11278 CCGCCATCTCAAAAACTTCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11337
Qy 6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
Db 11338 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11397
Qy 6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
Db 11398 GACCGCTACGCCCCAATGACCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11457
Qy 6242 GCATAATGCCACGCGCCGCT----------TGACACT 6268

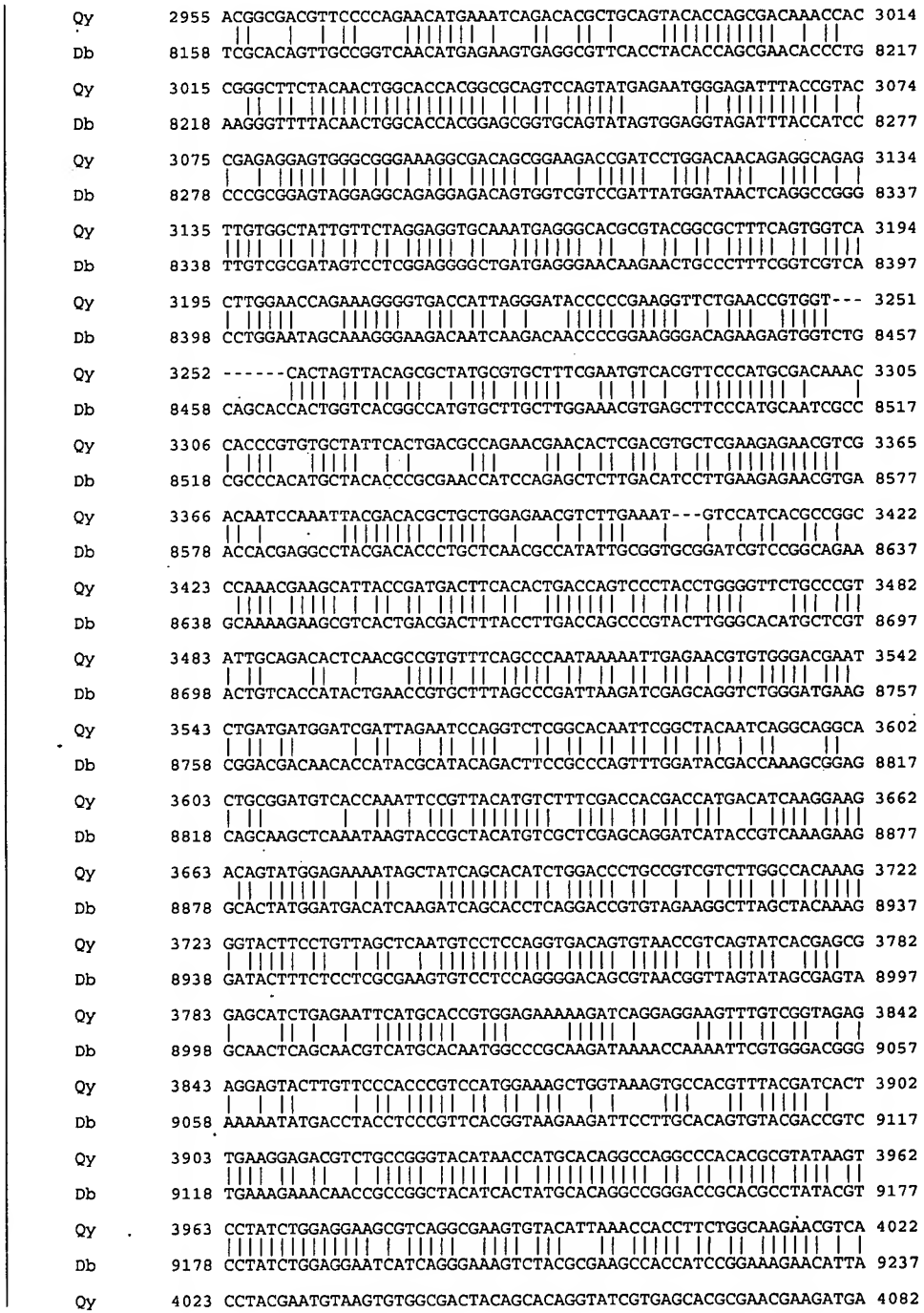
```
Db      11458 GCATAATGCATCAGGCTGGTATATTAGATCCCCGCTTACCGCGGGCAATATAGCAACACC 11517

Qy       6269 AAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGT 6318

Db      11518 AAAACTCGACGTATTTCCGAGGAAGCGCAGTGCATAATGCTGCGCAGTGT 11567


RESULT 13
US-09-102-248-4
; Sequence 4, Application US/09102248
; Patent No. 6008035
;  GENERAL INFORMATION:
;    APPLICANT:  Johnston, Robert E.
;    APPLICANT:  Davis, Nancy L.
;    APPLICANT:  Simpson, Dennis A.
;    TITLE OF INVENTION:  System for the In Vivo Delivery and
;    TITLE OF INVENTION:  Expression of Heterologous Genes in the Bone Marrow
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson, P.A.
;      STREET:  1211 East Morehead Street
;      CITY:  Charlotte
;      STATE:  No. 6008035th Carolina
;      COUNTRY:  USA
;      ZIP:  28234
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/102,248
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/801,263
;      FILING DATE:  19-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sibley, Kenneth D.
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5470-147
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-420-2200
;      TELEFAX:  919-881-3175
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11717 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-102-248-4

  Query Match             36.8%;  Score 2386.4;  DB 3;  Length 11717;
  Best Local Similarity   64.5%;  Pred. No. 0;
  Matches 3745;  Conservative    0;  Mismatches 1991;  Indels   74;  Gaps    9;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639

Db       5761 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5820

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699

Db       5821 AAATGGAGTCCGTTCTGCAGAATCAGCTTACAGAACCGACCTTGGAGCGCAATGTTCTGG 5880

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759

Db       5881 AAAGAATCTACGCCCCGGTGCTCGACACGTCGAAAGAGGAACAGCTCAAACTCAGGTACC 5940

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819

Db       5941 AGATGATGCCCACCGAAGCCAACAAAAGCAGGTACCAGTCTAGAAAAGTAGAAAATCAGA 6000

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879

Db       6001 AAGCCATAACCACTGAGCGACTGCTTTCAGGGCTACGACTGTATAACTCTGCCACAG--- 6057

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939

Db       6058 ATCAGCCAGAATGCTATAAGATCACCTACCCGAAACCATCGTATTCCAGCAGTGTACCGG 6117

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999

Db       6118 CGAACTACTCTGACCCAAAGTTTGCTGTAGCTGTTTGCAACAACTATCTGCATGAGAATT 6177

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059

Db       6178 ACCCGACGGTAGCATCTTATCAGATCACCGACGAGTACGATGCTTACTTGGATATGGTAG 6237

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119

Db       6238 ACGGGACAGTCGCTTGCCTAGATACTGCAACTTTTTGCCCCGCCAAGCTTAGAAGTTACC 6297

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179

Db       6298 CGAAAAGACACGAGTATAGAGCCCCAAACACTCGCAGTGCGGTTCCATCAGCGATGCAGA 6357

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239

Db       6358 ACACGTTGCAAAACGTGCTCATTGCCGCGACTAAAAGAAACTGCAACGTCACACAAATGC 6417

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299

Db       6418 GTGAATTGCCAACACTGGACTCAGCGACATTCAACGTTGAATGCTTTCGAAAATATGCAT 6477

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359

Db       6478 GTAATGACGAGTATTGGGAGGAGTTTGCCCGAAAGCCAATTAGGATCACTACTGAGTTCG 6537

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419

Db       6538 TTACCGCATACGTGGCCAGACTGAAAGGCCCTAAGGCCGCCGCACTGTTCGCAAAGACGC 6597

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479

Db       6598 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGGTTCGTCATGGACATGAAAAGAG 6657

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539

Db       6658 ACGTGAAAGTTACACCTGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGCTAC 6717

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599

Db       6718 AAGCCGCAGAACCCCTGGCGACCGCTTACCTGTGCGGGATCCACCGGGAGTTAGTGCGCA 6777

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659

Db       6778 GGCTTACAGCCGTCTTGCTACCCAACATTCACACGCTTTTTGACATGTCGGCGGAGGACT 6837

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719

Db       6838 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGTGACCCGGTACTGGAGACGGATATCG 6897

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779

Db       6898 CCTCGTTCGACAAAAGCCAAGACGACGCTATGGCGTTAACTGGCCTGATGATCTTGGAAG 6957

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839

Db       6958 ACCTGGGTGTGGACCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 7017

Qy       1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899

Db       7018 CCACCCATCTGCCCACGGGTACCCGTTTCAAATTCGGGGCGATGATGAAATCCGGAATGT 7077
```

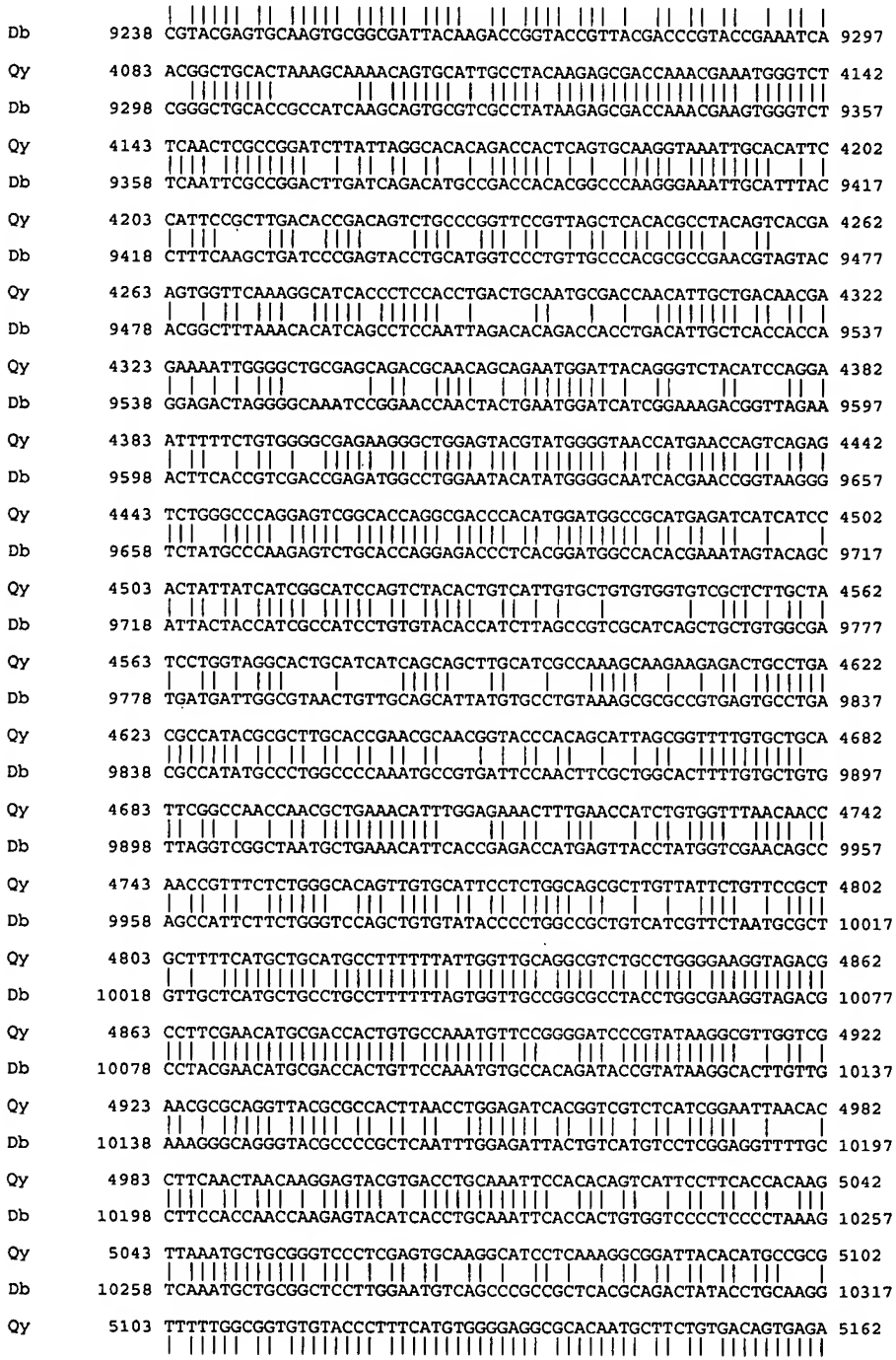

Qy 1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
Db 7078 TCCTCACGCTCTTTGTCAACACAGTTCTGAATGTCGTTATCGCCAGCAGAGTATTGGAGG 7137
Qy 1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
Db 7138 AGCGGCTTAAAACGTCCAAATGTGCAGCATTTATCGGCGACGACAACATCATACACGGAG 7197
Qy 2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
Db 7198 TAGTATCTGACAAAGAAATGGCTGAGAGGTGTGCCACCTGGCTCAACATGGAGGTTAAGA 7257
Qy 2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139
Db 7258 TCATTGACGCAGTCATCGGCGAGAGACCGCCTTACTTCTGCGGTGGATTCATCTTGCAAG 7317
Qy 2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199
Db 7318 ATTCGGTTACCTCCACAGCGTGTCGCGTGGCGGACCCCTTGAAAAGGCTGTTTAAGTTGG 7377
Qy 2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
Db 7378 GTAAACCGCTCCCAGCCGACGACGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 7437
Qy 2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319
Db 7438 AAACAAAGGCGTGGTTTAGAGTAGGTATAACAGACACCTTAGCAGTGGCCGTGGCAACTC 7497
Qy 2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379
Db 7498 GGTATGAGGTAGACAACATCACACCTGTCCTGCTGGCATTGAGAACTTTTGCCCAGAGCA 7557
Qy 2380 TTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGG 2439
Db 7558 AAAGAGCATTTCAAGCCATCAGAGGGGAAATAAAGCATCTCTACGGTGGTCCTAAATAGT 7617
Qy 2440 TGACGTAGTA----------GACACGCACCTACCCACCGGCAGAATGTTTCCATACC 2486
Db 7618 CAGCATAGCACATTTCATCTGACTAATACCACAACACCACCACCATGAATAGAGGATTCT 7677
Qy 2487 CTCAGCTGAACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACC 2546
Db 7678 TTAACATGCTCGGCCGCCGCCCCTTCCCGGCCCCCACTGCCATGTGGAGGCCGCGGAGAA 7737
Qy 2547 CTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTA 2606
Db 7738 GGAGGCAGGCGGCCCCGATGCCTGCCCGCAATGGGCTGGCTTCCCAAATCCAGCAACTGA 7797
Qy 2607 GGAGGTCGATAGTCAACTTGACTTTCAAACA------ACGATCACCTAATCCGCCGCCAG 2660
Db 7798 CCACAGCCGTCAGTGCCCTAGTCATTGGACAGGCAACTAGACCTCAAACCCCACGCCCAC 7857
Qy 2661 GTCCACCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGA 2720
Db 7858 GCCCGCCGCCGCGCCAGAAGAAGCAGGCGCCAAAGCAACCACCGAAGCCGAAGAAACCAA 7917
Qy 2721 AGAAGCAGCAAGCCAAGAGGACGAAAC---GCAAGCCTAAACCAGGGAAACGACAACGTA 2777
Db 7918 AAACACAGGAGAAGAAGAAGAAGCAACCTGCAAAACCCAAACCCGGAAAGAGACAACGTA 7977
Qy 2778 TGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATC---ATGCTGAACGGCCAAGTGA 2834
Db 7978 TGGCACTCAAGTTGGAGGCCGACAGACTGTTCGACGTCAAAAATGAGGACGGAGATGTCA 8037
Qy 2835 ATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACTCCACGTTGAAGGAAAAA 2894
Db 8038 TCGGGCACGCACTGGCCATGGAAGGAAAGGTAATGAAACCACTCCACGTGAAAGGAACTA 8097
Qy 2895 TTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAGCATGTACGACTTGGAGT 2954
Db 8098 TTGACCACCCTGTGCTATCAAAGCTCAAATTCACCAAGTCGTCAGCATACGACATGGAGT 8157
Qy 2955 ACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTACACCAGCGACAAACCAC 3014
Db 8158 TCGCACAGTTGCCGGTCAACATGAGAAGTGAGGCGTTCACCTACACCAGCGAACACCCTG 8217
Qy 3015 CGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAATGGGAGATTTACCGTAC 3074
Db 8218 AAGGGTTTTACAACTGGCACCACGGAGCGGTGCAGTATAGTGGAGGTAGATTTACCATCC 8277
Qy 3075 CGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGACAACAGAGGCAGAG 3134
Db 8278 CCCGCGGAGTAGGAGGCAGAGGAGACAGTGGTCGTCCGATTATGGATAACTCAGGCCGGG 8337
Qy 3135 TTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGCGCTTTCAGTGGTCA 3194
Db 8338 TTGTCGCGATAGTCCTCGGAGGGGCTGATGAGGGAACAAGAACTGCCCTTTCGGTCGTCA 8397
Qy 3195 CTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTCTGAACCGTGGT--- 3251
Db 8398 CCTGGAATAGCAAAGGGAAGACAATCAAGACAACCCCGGAAGGGACAGAAGAGTGGTCTG 8457
Qy 3252 ------CACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGACAAAC 3305
Db 8458 CAGCACCACTGGTCACGGCCATGTGCTTGCTTGGAAACGTGAGCTTCCCATGCAATCGCC 8517
Qy 3306 CACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAACGTCG 3365
Db 8518 CGCCCACATGCTACACCCGCGAACCATCCAGAGCTCTTGACATCCTTGAAGAGAACGTGA 8577
Qy 3366 ACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAAT---GTCCATCACGCCGGC 3422
Db 8578 ACCACGAGGCCTACGACACCCTGCTCAACGCCATATTGCGGTGCGGATCGTCCGGCAGAA 8637
Qy 3423 CCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCCCGT 3482
Db 8638 GCAAAAGAAGCGTCACTGACGACTTTACCTTGACCAGCCCGTACTTGGGCACATGCTCGT 8697
Qy 3483 ATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACGAAT 3542
Db 8698 ACTGTCACCATACTGAACCGTGCTTTAGCCCGATTAAGATCGAGCAGGTCTGGGATGAAG 8757
Qy 3543 CTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAGGCA 3602
Db 8758 CGGACGACAACACCATACGCATACAGACTTCCGCCCAGTTTGGATACGACCAAAGCGGAG 8817
Qy 3603 CTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAAGGAAG 3662
Db 8818 CAGCAAGCTCAAATAAGTACCGCTACATGTCGCTCGAGCAGGATCATACCGTCAAAGAAG 8877
Qy 3663 ACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCACAAAG 3722
Db 8878 GCACTATGGATGACATCAAGATCAGCACCTCAGGACCGTGTAGAAGGCTTAGCTACAAAG 8937
Qy 3723 GGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCACGAGCG 3782
Db 8938 GATACTTTCTCCTCGCGAAGTGTCCTCCAGGGGACAGCGTAACGGTTAGTATAGCGAGTA 8997
Qy 3783 GAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTAGAG 3842
Db 8998 GCAACTCAGCAACGTCATGCACAATGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGG 9057
Qy 3843 AGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATCACT 3902
Db 9058 AAAAATATGACCTACCTCCCGTTCACGGTAAGAAGATTCCTTGCACAGTGTACGACCGTC 9117
Qy 3903 TGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAAGT 3962
Db 9118 TGAAAGAAACAACCGCCGGCTACATCACTATGCACAGGCCGGGACCGCACGCCTATACGT 9177
Qy 3963 CCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGTCA 4022
Db 9178 CCTATCTGGAGGAATCATCAGGGAAAGTCTACGCGAAGCCACCATCCGGAAAGAACATTA 9237
Qy 4023 CCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGATGA 4082


Db 9238 CGTACGAGTGCAAGTGCGGCGATTACAAGACCGGTACCGTTACGACCCGTACCGAAATCA 9297
Qy 4083 ACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGTCT 4142
Db 9298 CGGGCTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAACGAAGTGGGTCT 9357
Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9358 TCAATTCGCCGGACTTGATCAGACATGCCGACCACACGGCCCAAGGGAAATTGCATTTAC 9417
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9418 CTTTCAAGCTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAACGTAGTAC 9477
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9478 ACGGCTTTAAACACATCAGCCTCCAATTAGACACAGACCACCTGACATTGCTCACCACCA 9537
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9538 GGAGACTAGGGGCAAATCCGGAACCAACTACTGAATGGATCATCGGAAAGACGGTTAGAA 9597
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9598 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGCAATCACGAACCGGTAAGGG 9657
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db 9658 TCTATGCCCAAGAGTCTGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9717
Qy 4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db 9718 ATTACTACCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTGCTGTGGCGA 9777
Qy 4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db 9778 TGATGATTGGCGTAACTGTTGCAGCATTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9837
Qy 4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db 9838 CGCCATATGCCCTGGCCCCAAATGCCGTGATTCCAACTTCGCTGGCACTTTTGTGCTGTG 9897
Qy 4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db 9898 TTAGGTCGGCTAATGCTGAAACATTCACCGAGACCATGAGTTACCTATGGTCGAACAGCC 9957
Qy 4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
Db 9958 AGCCATTCTTCTGGGTCCAGCTGTGTATACCCCTGGCCGCTGTCATCGTTCTAATGCGCT 10017
Qy 4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
Db 10018 GTTGCTCATGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10077
Qy 4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
Db 10078 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10137
Qy 4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db 10138 AAAGGGCAGGGTACGCCCCGCTCAATTTGGAGATTACTGTCATGTCCTCGGAGGTTTTGC 10197
Qy 4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db 10198 CTTCCACCAACCAAGAGTACATCACCTGCAAATTCACCACTGTGGTCCCCTCCCCTAAAG 10257
Qy 5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db 10258 TCAAATGCTGCGGCTCCTTGGAATGTCAGCCCGCCGCTCACGCAGACTATACCTGCAAGG 10317
Qy 5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db 10318 TCTTTGGAGGGGTGTACCCCTTCATGTGGGGAGGAGCACAATGTTTTTGCGACAGTGAGA 10377
Qy 5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db 10378 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGCAGATTGCGCGACTGACCACGCGC 10437
Qy 5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db 10438 AGGCGATTAAGGTGCATACTGCCGCGATGAAAGTAGGACTACGTATAGTGTACGGGAACA 10497
Qy 5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db 10498 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACCTGA 10557
Qy 5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db 10558 AAGTCATAGCTGGACCAATTTCAGCATCGTTTACACCATTCGATCACAAGGTCGTTATCC 10617
Qy 5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db 10618 ATCGCGGCCTGGTGTACAACTATGACTTCCCGGAATACGGAGCGATGAAACCAGGAGCGT 10677
Qy 5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db 10678 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAAGATCTCATCGCCAGCACAGACATTA 10737
Qy 5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db 10738 GACTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCTGGAT 10797
Qy 5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db 10798 TCGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCCCCTTTCGGGTGCA 10857
Qy 5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db 10858 AGATTGCAGTCAATCCGCTTCGAGCGGTGGACTGCTCATACGGGAACATTCCCATCTCTA 10917
Qy 5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db 10918 TCGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10977
Qy 5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db 10978 AATGTGATGTCAGTGAGTGCACTTACTCAGCGGACTTCGGCGGGATGGCTACCCTGCAGT 11037
Qy 5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db 11038 ATGTATCCGACCGCGAAGGACAATGCCCTGTACATTCGCATTCGAGCACAGCAACCCTCC 11097
Qy 5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db 11098 AAGAGTCGACAGTTCATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTCAGCACCGCGA 11157
Qy 5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db 11158 GCCCACAGGCGAACTTTATTGTATCGCTGTGTGGTAAGAAGACAACATGCAATGCAGAAT 11217
Qy 6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db 11218 GCAAACCACCAGCTGACCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTCCAAG 11277
Qy 6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db 11278 CCGCCATCTCAAAAACTTCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11337
Qy 6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
Db 11338 TATTAATTATAGGACTTATGATTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11397
Qy 6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
Db 11398 GACCGCTACGCCCCAATGACCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11457

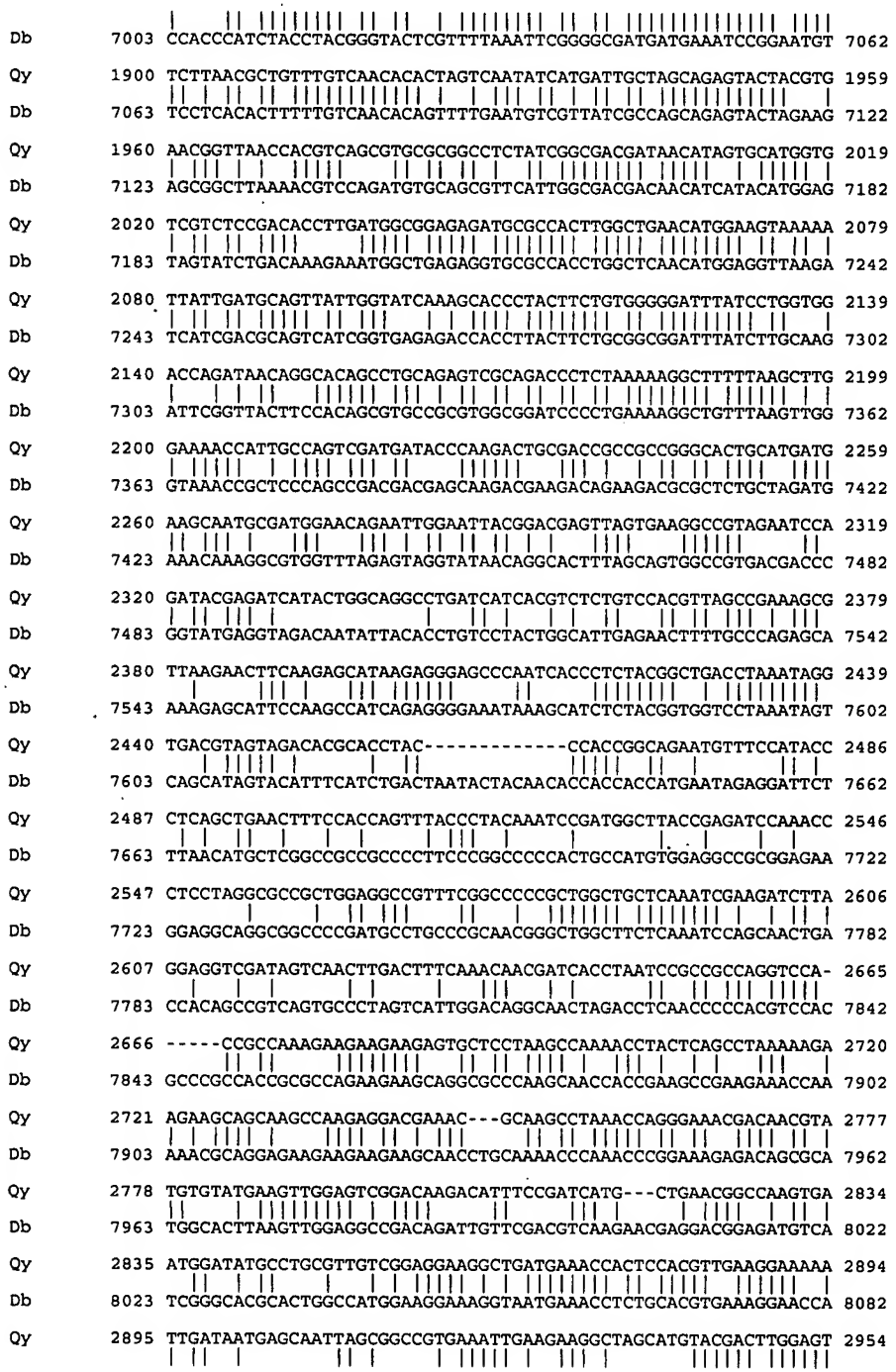
```
Qy       6242 GCATAATGCCACGCGCCGCT--------------------------------TGACACT 6268
Db      11458 GCATAATGCATCAGGCTGGTATATTAGATCCCCGCTTACCGCGGGCAATATAGCAACACC 11517

Qy       6269 AAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGT 6318
Db      11518 AAAACTCGACGTATTTCCGAGGAAGCGCAGTGCATAATGCTGCGCAGTGT 11567


RESULT 14
US-09-367-764-4
; Sequence 4, Application US/09367764
; Patent No. 6583121
;  GENERAL INFORMATION:
;    APPLICANT:  Johnston, Robert E.
;    APPLICANT:  Davis, Nancy L.
;    APPLICANT:  Simpson, Dennis A.
;    TITLE OF INVENTION:  System for the In Vivo Delivery and
;    TITLE OF INVENTION:  Expression of Heterologous Genes in the Bone Marrow
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson, P.A.
;      STREET:  1211 East Morehead Street
;      CITY:  Charlotte
;      STATE:  No. 6583121th Carolina
;      COUNTRY:  USA
;      ZIP:  28234
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/367,764
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/801,263
;      FILING DATE:  19-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sibley, Kenneth D.
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5470-147
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-420-2200
;      TELEFAX:  919-881-3175
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11717 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-367-764-4

  Query Match             36.8%;  Score 2386.4;  DB 4;  Length 11717;
  Best Local Similarity   64.5%;  Pred. No. 0;
  Matches 3745;  Conservative    0;  Mismatches 1991;  Indels   74;  Gaps    9;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
Db       5761 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5820

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
Db       5821 AAATGGAGTCCGTTCTGCAGAATCAGCTTACAGAACCGACCTTGGAGCGCAATGTTCTGG 5880

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
Db       5881 AAAGAATCTACGCCCCGGTGCTCGACACGTCGAAAGAGGAACAGCTCAAACTCAGGTACC 5940

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
Db       5941 AGATGATGCCCACCGAAGCCAACAAAAGCAGGTACCAGTCTAGAAAAGTAGAAAATCAGA 6000

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
Db       6001 AAGCCATAACCACTGAGCGACTGCTTTCAGGGCTACGACTGTATAACTCTGCCACAG--- 6057

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
Db       6058 ATCAGCCAGAATGCTATAAGATCACCTACCCGAAACCATCGTATTCCAGCAGTGTACCGG 6117

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
Db       6118 CGAACTACTCTGACCCAAAGTTTGCTGTAGCTGTTTGCAACAACTATCTGCATGAGAATT 6177

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
Db       6178 ACCCGACGGTAGCATCTTATCAGATCACCGACGAGTACGATGCTTACTTGGATATGGTAG 6237

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
Db       6238 ACGGGACAGTCGCTTGCCTAGATACTGCAACTTTTTGCCCCGCCAAGCTTAGAAGTTACC 6297

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
Db       6298 CGAAAAGACACGAGTATAGAGCCCCAAACACTCGCAGTGCGGTTCCATCAGCGATGCAGA 6357

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
Db       6358 ACACGTTGCAAAACGTGCTCATTGCCGCGACTAAAAGAAACTGCAACGTCACACAAATGC 6417

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
Db       6418 GTGAATTGCCAACACTGGACTCAGCGACATTCAACGTTGAATGCTTTCGAAAATATGCAT 6477

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
Db       6478 GTAATGACGAGTATTGGGAGGAGTTTGCCCGAAAGCCAATTAGGATCACTACTGAGTTCG 6537

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
Db       6538 TTACCGCATACGTGGCCAGACTGAAAGGCCCTAAGGCCGCCGCACTGTTCGCAAAGACGC 6597

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
Db       6598 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGGTTCGTCATGGACATGAAAAGAG 6657

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
Db       6658 ACGTGAAAGTTACACCTGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGCTAC 6717

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
Db       6718 AAGCCGCAGAACCCCTGGCGACCGCTTACCTGTGCGGGATCCACCGGGAGTTAGTGCGCA 6777

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
Db       6778 GGCTTACAGCCGTCTTGCTACCCAACATTCACACGCTTTTTGACATGTCGGCGGAGGACT 6837

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
Db       6838 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGTGACCCGGTACTGGAGACGGATATCG 6897

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
Db       6898 CCTCGTTCGACAAAAGCCAAGACGACGCTATGGCGTTAACTGGCCTGATGATCTTGGAAG 6957

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
Db       6958 ACCTGGGTGTGGACCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 7017

Qy       1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
Db       7018 CCACCCATCTGCCCACGGGTACCCGTTTCAAATTCGGGGCGATGATGAAATCCGGAATGT 7077
```

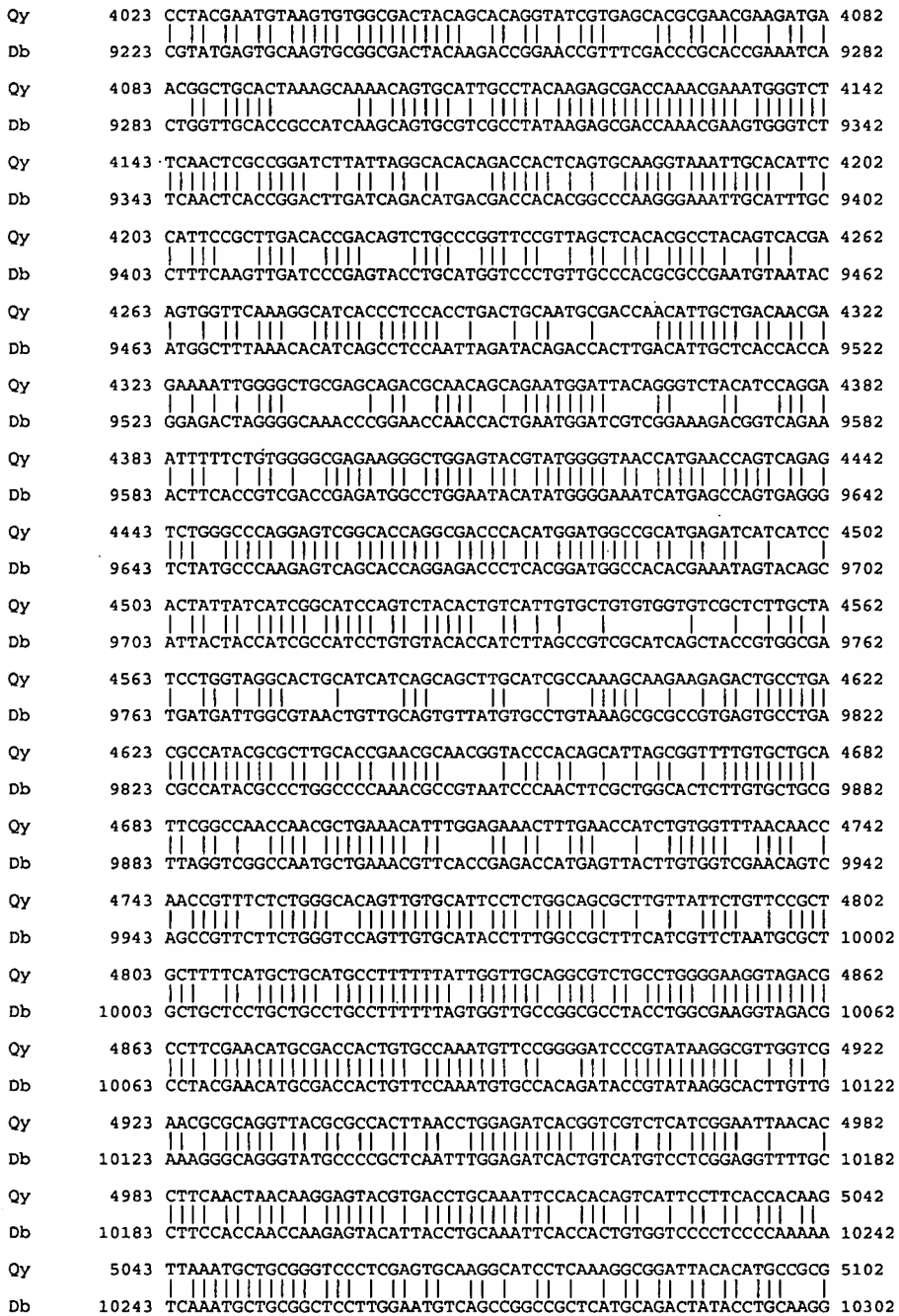

Qy 1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
Db 7078 TCCTCACGCTCTTTGTCAACACAGTTCTGAATGTCGTTATCGCCAGCAGAGTATTGGAGG 7137
Qy 1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
Db 7138 AGCGGCTTAAAACGTCCAAATGTGCAGCATTTATCGGCGACGACAACATCATACACGGAG 7197
Qy 2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
Db 7198 TAGTATCTGACAAAGAAATGGCTGAGAGGTGTGCCACCTGGCTCAACATGGAGGTTAAGA 7257
Qy 2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139
Db 7258 TCATTGACGCAGTCATCGGCGAGAGACCGCCTTACTTCTGCGGTGGATTCATCTTGCAAG 7317
Qy 2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199
Db 7318 ATTCGGTTACCTCCACAGCGTGTCGCGTGGCGGACCCCTTGAAAAGGCTGTTTAAGTTGG 7377
Qy 2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
Db 7378 GTAAACCGCTCCCAGCCGACGACGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 7437
Qy 2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319
Db 7438 AAACAAAGGCGTGGTTTAGAGTAGGTATAACAGACACCTTAGCAGTGGCCGTGGCAACTC 7497
Qy 2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379
Db 7498 GGTATGAGGTAGACAACATCACACCTGTCCTGCTGGCATTGAGAACTTTTGCCCAGAGCA 7557
Qy 2380 TTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGG 2439
Db 7558 AAAGAGCATTTCAAGCCATCAGAGGGGAAATAAAGCATCTCTACGGTGGTCCTAAATAGT 7617
Qy 2440 TGACGTAGTA----------GACACGCACCTACCCACCGGCAGAATGTTTCCATACC 2486
Db 7618 CAGCATAGCACATTTCATCTGACTAATACCACAACACCACCACCATGAATAGAGGATTCT 7677
Qy 2487 CTCAGCTGAACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACC 2546
Db 7678 TTAACATGCTCGGCCGCCGCCCCTTCCCGGCCCCCACTGCCATGTGGAGGCCGCGGAGAA 7737
Qy 2547 CTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTA 2606
Db 7738 GGAGGCAGGCGGCCCCGATGCCTGCCCGCAATGGGCTGGCTTCCCAAATCCAGCAACTGA 7797
Qy 2607 GGAGGTCGATAGTCAACTTGACTTTCAAACA------ACGATCACCTAATCCGCCGCCAG 2660
Db 7798 CCACAGCCGTCAGTGCCCTAGTCATTGGACAGGCAACTAGACCTCAAACCCCACGCCCAC 7857
Qy 2661 GTCCACCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGA 2720
Db 7858 GCCCGCCGCCGCGCCAGAAGAAGCAGGCGCCAAAGCAACCACCGAAGCCGAAGAAACCAA 7917
Qy 2721 AGAAGCAGCAAGCCAAGAGGACGAAAC---GCAAGCCTAAACCAGGGAAACGACAACGTA 2777
Db 7918 AAACACAGGAGAAGAAGAAGAAGCAACCTGCAAAACCCAAACCCGGAAAGAGACAACGTA 7977
Qy 2778 TGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATC---ATGCTGAACGGCCAAGTGA 2834
Db 7978 TGGCACTCAAGTTGGAGGCCGACAGACTGTTCGACGTCAAAAATGAGGACGGAGATGTCA 8037
Qy 2835 ATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACTCCACGTTGAAGGAAAAA 2894
Db 8038 TCGGGCACGCACTGGCCATGGAAGGAAAGGTAATGAAACCACTCCACGTGAAAGGAACTA 8097
Qy 2895 TTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAGCATGTACGACTTGGAGT 2954
Db 8098 TTGACCACCCTGTGCTATCAAAGCTCAAATTCACCAAGTCGTCAGCATACGACATGGAGT 8157


Qy 2955 ACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTACACCAGCGACAAACCAC 3014
Db 8158 TCGCACAGTTGCCGGTCAACATGAGAAGTGAGGCGTTCACCTACACCAGCGAACACCCTG 8217
Qy 3015 CGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAATGGGAGATTTACCGTAC 3074
Db 8218 AAGGGTTTTACAACTGGCACCACGGAGCGGTGCAGTATAGTGGAGGTAGATTTACCATCC 8277
Qy 3075 CGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGACAACAGAGGCAGAG 3134
Db 8278 CCCGCGGAGTAGGAGGCAGAGGAGACAGTGGTCGTCCGATTATGGATAACTCAGGCCGGG 8337
Qy 3135 TTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGCGCTTTCAGTGGTCA 3194
Db 8338 TTGTCGCGATAGTCCTCGGAGGGGCTGATGAGGGAACAAGAACTGCCCTTTCGGTCGTCA 8397
Qy 3195 CTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTCTGAACCGTGGT--- 3251
Db 8398 CCTGGAATAGCAAAGGGAAGACAATCAAGACAACCCCGGAAGGGACAGAAGAGTGGTCTG 8457
Qy 3252 ------CACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGACAAAC 3305
Db 8458 CAGCACCACTGGTCACGGCCATGTGCTTGCTTGGAAACGTGAGCTTCCCATGCAATCGCC 8517
Qy 3306 CACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAACGTCG 3365
Db 8518 CGCCCACATGCTACACCCGCGAACCATCCAGAGCTCTTGACATCCTTGAAGAGAACGTGA 8577
Qy 3366 ACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAAT---GTCCATCACGCCGGC 3422
Db 8578 ACCACGAGGCCTACGACACCCTGCTCAACGCCATATTGCGGTGCGGATCGTCCGGCAGAA 8637
Qy 3423 CCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCCCGT 3482
Db 8638 GCAAAAGAAGCGTCACTGACGACTTTACCTTGACCAGCCCGTACTTGGGCACATGCTCGT 8697
Qy 3483 ATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACGAAT 3542
Db 8698 ACTGTCACCATACTGAACCGTGCTTTAGCCCGATTAAGATCGAGCAGGTCTGGGATGAAG 8757
Qy 3543 CTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAGGCA 3602
Db 8758 CGGACGACAACACCATACGCATACAGACTTCCGCCCAGTTTGGATACGACCAAAGCGGAG 8817
Qy 3603 CTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAAGGAAG 3662
Db 8818 CAGCAAGCTCAAATAAGTACCGCTACATGTCGCTCGAGCAGGATCATACCGTCAAAGAAG 8877
Qy 3663 ACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCACAAAG 3722
Db 8878 GCACTATGGATGACATCAAGATCAGCACCTCAGGACCGTGTAGAAGGCTTAGCTACAAAG 8937
Qy 3723 GGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCACGAGCG 3782
Db 8938 GATACTTTCTCCTCGCGAAGTGTCCTCCAGGGGACAGCGTAACGGTTAGTATAGCGAGTA 8997
Qy 3783 GAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTAGAG 3842
Db 8998 GCAACTCAGCAACGTCATGCACAATGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGG 9057
Qy 3843 AGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATCACT 3902
Db 9058 AAAAATATGACCTACCTCCCGTTCACGGTAAGAAGATTCCTTGCACAGTGTACGACCGTC 9117
Qy 3903 TGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAAGT 3962
Db 9118 TGAAAGAAACAACCGCCGGCTACATCACTATGCACAGGCCGGGACCGCACGCCTATACGT 9177
Qy 3963 CCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGTCA 4022
Db 9178 CCTATCTGGAGGAATCATCAGGGAAAGTCTACGCGAAGCCACCATCCGGAAAGAACATTA 9237
Qy 4023 CCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGATGA 4082


Db 9238 CGTACGAGTGCAAGTGCGGCGATTACAAGACCGGTACCGTTACGACCCGTACCGAAATCA 9297
Qy 4083 ACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGTCT 4142
Db 9298 CGGGCTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAACGAAGTGGGTCT 9357
Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9358 TCAATTCGCCGGACTTGATCAGACATGCCGACCACACGGCCCAAGGGAAATTGCATTTAC 9417
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9418 CTTTCAAGCTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAACGTAGTAC 9477
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9478 ACGGCTTTAAACACATCAGCCTCCAATTAGACACAGACCACCTGACATTGCTCACCACCA 9537
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9538 GGAGACTAGGGGCAAATCCGGAACCAACTACTGAATGGATCATCGGAAAGACGGTTAGAA 9597
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9598 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGCAATCACGAACCGGTAAGGG 9657
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db 9658 TCTATGCCCAAGAGTCTGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9717
Qy 4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db 9718 ATTACTACCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTGCTGTGGCGA 9777
Qy 4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db 9778 TGATGATTGGCGTAACTGTTGCAGCATTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9837
Qy 4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db 9838 CGCCATATGCCCTGGCCCCAAATGCCGTGATTCCAACTTCGCTGGCACTTTTGTGCTGTG 9897
Qy 4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db 9898 TTAGGTCGGCTAATGCTGAAACATTCACCGAGACCATGAGTTACCTATGGTCGAACAGCC 9957
Qy 4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
Db 9958 AGCCATTCTTCTGGGTCCAGCTGTGTATACCCCTGGCCGCTGTCATCGTTCTAATGCGCT 10017
Qy 4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
Db 10018 GTTGCTCATGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10077
Qy 4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
Db 10078 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10137
Qy 4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db 10138 AAAGGGCAGGGTACGCCCCGCTCAATTTGGAGATTACTGTCATGTCCTCGGAGGTTTTGC 10197
Qy 4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db 10198 CTTCCACCAACCAAGAGTACATCACCTGCAAATTCACCACTGTGGTCCCCTCCCCTAAAG 10257
Qy 5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db 10258 TCAAATGCTGCGGCTCCTTGGAATGTCAGCCCGCCGCTCACGCAGACTATACCTGCAAGG 10317
Qy 5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db 10318 TCTTTGGAGGGGTGTACCCCTTCATGTGGGGAGGAGCACAATGTTTTTGCGACAGTGAGA 10377
Qy 5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db 10378 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGCAGATTGCGCGACTGACCACGCGC 10437
Qy 5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db 10438 AGGCGATTAAGGTGCATACTGCCGCGATGAAAGTAGGACTACGTATAGTGTACGGGAACA 10497
Qy 5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db 10498 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACCTGA 10557
Qy 5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db 10558 AAGTCATAGCTGGACCAATTTCAGCATCGTTTACACCATTCGATCACAAGGTCGTTATCC 10617
Qy 5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db 10618 ATCGCGGCCTGGTGTACAACTATGACTTCCCGGAATACGGAGCGATGAAACCAGGAGCGT 10677
Qy 5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db 10678 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAAGATCTCATCGCCAGCACAGACATTA 10737
Qy 5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db 10738 GACTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCGCATCTGGAT 10797
Qy 5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db 10798 TCGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCCCCTTTCGGGTGCA 10857
Qy 5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db 10858 AGATTGCAGTCAATCCGCTTCGAGCGGTGGACTGCTCATACGGGAACATTCCCATCTCTA 10917
Qy 5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db 10918 TCGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10977
Qy 5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db 10978 AATGTGATGTCAGTGAGTGCACTTACTCAGCGGACTTCGGCGGGATGGCTACCCTGCAGT 11037
Qy 5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db 11038 ATGTATCCGACCGCGAAGGACAATGCCCTGTACATTCGCATTCGAGCACAGCAACCCTCC 11097
Qy 5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db 11098 AAGAGTCGACAGTTCATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTCAGCACCGCGA 11157
Qy 5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db 11158 GCCCACAGGCGAACTTTATTGTATCGCTGTGTGGTAAGAAGACAACATGCAATGCAGAAT 11217
Qy 6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db 11218 GCAAACCACCAGCTGACCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTCCAAG 11277
Qy 6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db 11278 CCGCCATCTCAAAAACTTCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11337
Qy 6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
Db 11338 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11397
Qy 6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241
Db 11398 GACCGCTACGCCCCAATGACCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11457

```
Qy       6242 GCATAATGCCACGCGCCGCT--------------------------------TGACACT 6268
              |||||||||  |  || | |                                ||||
Db      11458 GCATAATGCATCAGGCTGGTATATTAGATCCCCGCTTACCGCGGGCAATATAGCAACACC 11517

Qy       6269 AAAACTCGATGTATTTCCGAGGAAGCACAGTGCATAATGCTGTGCAGTGT 6318
              ||||||||| |||||||||||||||| |||||||||||||| |||||||
Db      11518 AAAACTCGACGTATTTCCGAGGAAGCGCAGTGCATAATGCTGCGCAGTGT 11567


RESULT 15
US-09-415-784-103
; Sequence 103, Application US/09415784
; Patent No. 6391632
;  GENERAL INFORMATION:
;       APPLICANT: Dubensky Jr., Thomas W.
;                  Polo, John M.
;                  Belli, Barbara A.
;                  Schlesinger, Sondra
;                  Dryga, Sergey A.
;                  Frolov, Ilya
;       TITLE OF INVENTION: RECOMBINANT ALPHAVIRUS-BASED VECTORS
;                           WITH REDUCED INHIBITION OF CELLULAR MACRO-MOLECULAR
;                           SYNTHESIS
;       NUMBER OF SEQUENCES: 125
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Seed Intellectual Property Law Group PLLC
;            STREET: 701 Fifth Avenue, Suite 6300
;            CITY: Seattle
;            STATE: Washington
;            COUNTRY: USA
;            ZIP: 98104-7092
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/415,784
;            FILING DATE: 08-Oct-1999
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: McMasters, David D.
;            REGISTRATION NUMBER: 33,963
;            REFERENCE/DOCKET NUMBER: 930049.457D1 /1196.006
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (206) 622-4900
;            TELEFAX: (206) 682-6031
;  INFORMATION FOR SEQ ID NO: 103:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 11740 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 103:
US-09-415-784-103

  Query Match           36.7%;  Score 2380.4;  DB 4;  Length 11740;
  Best Local Similarity  64.6%;  Pred. No. 0;
  Matches 3693;  Conservative    0;  Mismatches 1986;  Indels   41;  Gaps    8;

Qy        580 ACTGACGGTATGAAGCGGGAGCGTATATTTTCTCATCGGAAACAGGCCAAGGTCACCTTC 639
              ||||||     |  |  || | ||| || || ||  |||| |||||||  || ||| | |
Db       5746 ACTGACTAACCGGGGTAGGTGGGTACATATTTTCGACGGACACAGGCCCTGGGCACTTGC 5805

Qy        640 AACAGAAATCAGTACGTCAATGTAAACTACAAGAACCTATATTGGATCGGGCCGTCCATG 699
              || |||| || || |  ||     | ||   |||||| |  ||||| ||    ||||  |
Db       5806 AAAAGAAGTCCGTTCTGCAGAACCAGCTTACAGAACCGACCTTGGAGCGCAATGTCCTGG 5865

Qy        700 AGAAGTATTACGCCCCGCGCCTCGATCTCGAAAGAGAGAAAATGTTACAGAAGAAACTGC 759
              | |    | | ||||||   |||||       | |||| ||    |  |    |     |
Db       5866 AAAGAATTCATGCCCCGGTGCTCGACACGTCGAAAGAGGAACAACTCAAACTCAGGTACC 5925

Qy        760 AATTATGCGCCTCTGAAGGAAATAGAAGCAGGTATCAATCACGAAAAGTAGAAAATATGA 819
              |  |      || | |||| || | ||| ||||| || || || ||||||||||||  ||
Db       5926 AGATGATGCCCACCGAAGCCAACAAAAGTAGGTACCAGTCTCGTAAAGTAGAAAATCAGA 5985

Qy        820 AAGCAATTACAGCGGAGCGACTCATTTCTGGATTGGGCACATATCTATCATCAGAAGTGA 879
              |||| || ||  | |||||||| | || ||| |  |    |||   ||  |   ||
Db       5986 AAGCCATAACCACTGAGCGACTACTGTCAGGACTACGACTGTATAACTCTGCCACAG--- 6042

Qy        880 ATCCTGTCGAGTGTTACAGAGTCAATTATCCTGTACCAATCTACTCGTCAACGGTAATTA 939
              |||      || || || |   ||| |||||  |||| | |||||    |  |||
Db       6043 ATCAGCCAGAATGCTATAAGATCACCTATCCGAAACCATTGTACTCCAGTAGCGTACCGG 6102

Qy        940 ACAGGTTTACATCTGCAGAGGTCGCGGTTAAAACGTGCAACTTAGTTATCCAAGAGAATT 999
              |  |   |   | || || |||| ||        || |||     | | || ||||| |
Db       6103 CGAACTACTCCGATCCACAGTTCGCTGTAGCTGTCTGTAACAACTATCTGCATGAGAACT 6162

Qy       1000 ACCCTACAGTAGCCAGTTATTGTATAACAGATGAATACGATGCGTATCTTGACATGGTGG 1059
              | || ||||||||   ||||   || || || || |||||||| ||  | ||| ||||| |
Db       6163 ATCCGACAGTAGCATCTTATCAGATTACTGACGAGTACGATGCTTACTTGGATATGGTAG 6222

Qy       1060 ACGGCGCATCGTGCTGTCTAGATACAGCCACTTTTTGTCCGGCTAAACTGAGAAGCTACC 1119
              ||||  ||     ||| || ||||| || || || || || ||||| || ||||| ||||
Db       6223 ACGGGACAGTCGCCTGCCTGGATACTGCAACCTTCTGCCCCGCTAAGCTTAGAAGTTACC 6282

Qy       1120 CAAAGAAGCATAGCTATTTGCAGCCAGAGATAAGATCAGCCGTCCCATCGCCTATACAGA 1179
              | || || |||   |||      || | ||  |    || || ||||  | || ||||  |
Db       6283 CGAAAAAACATGAGTATAGAGCCCCGAATATCCGCAGTGCGGTTCCATCAGCGATGCAGA 6342

Qy       1180 ATACATTACAAAATGTATTGGCTGCAGCTACTAAAAGGAATTGCAACGTTACCCAAATGC 1239
              | ||  |||||||||  |   ||| || |||||||| |||||||||| || || ||||
Db       6343 ACACGCTACAAAATGTGCTCATTGCCGCAACTAAAAGAAATTGCAACGTCACGCAGATGC 6402

Qy       1240 GAGAATTACCTGTCTTAGATTCGGCGGCATTTAATGTTGATTGTTTCAAGAAATACGCAT 1299
              | ||| | ||      | || || ||| |||| ||||| || || ||    ||||| ||||
Db       6403 GTGAACTGCCAACACTGGACTCAGCGACATTCAATGTCGAATGCTTTCGAAAATATGCAT 6462

Qy       1300 GCAATGATGAGTACTGGGATACCTTTCGCGATAACCCTATTCGGCTAACTACAGAGAACG 1359
              | ||||| ||||| |||||    ||         || || ||| || | || || |||   |
Db       6463 GTAATGACGAGTATTGGGAGGAGTTCGCTCGGAAGCCAATTAGGATTACCACTGAGTTTG 6522

Qy       1360 TTACGCAATATGTGACAAAGCTGAAAGGGCCGAAAGCAGCAGCATTGTTTGCGAATACTC 1419
              | ||   ||||||  | |  |||||||| || || || || ||| | ||||| || ||
Db       6523 TCACCGCATATGTAGCTAGACTGAAAGGCCCTAAGGCCGCCGCACTATTTGCAAAGACGT 6582

Qy       1420 ATAATCTAAAACCGTTGCAGGAGATACCAATGGATCAATTCGTCATGGATCTAAAGAGAG 1479
              ||||| |   || ||||| ||  | || ||||||  |||||||||||  | || ||||
Db       6583 ATAATTTGGTCCCATTGCAAGAAGTGCCTATGGATAGATTCGTCATGGACATGAAAAGAG 6642

Qy       1480 ATGTCAAAGTTACTCCCGGCACGAAACATACAGAGGAGCGGCCTAAGGTGCAGGTTATTC 1539
              | || |||||||| || |||||||||| ||||| ||  | || || || || || | | |
Db       6643 ACGTGAAAGTTACACCAGGCACGAAACACACAGAAGAAAGACCGAAAGTACAAGTGATAC 6702

Qy       1540 AGGCTGCAGATCCCCTTGCTACCGCTTACCTTTGCGGGATCCATCGGGAATTAGTCCGTA 1599
              | || ||||| ||||| || || |||||| | |||||||| || ||||||||||| ||||
Db       6703 AAGCCGCAGAACCCCTGGCGACTGCTTACTTATGCGGGATTCACCGGGAATTAGTGCGTA 6762

Qy       1600 GACTGAATGCGGTGCTTCTGCCAAATATCCATACTCTCTTCGACATGTCAGCGGAAGATT 1659
              | || |  || ||  | || ||||| || || || || || |||||||| ||||| ||||
Db       6763 GGCTTACGGCCGTCTTGCTTCCAAACATTCACACGCTTTTTGACATGTCGGCGGAGGATT 6822

Qy       1660 TTGATGCGATTATTGCTGAACATTTCCACCACGGCGACCCAGTATTGGAAACGGACATCG 1719
              ||||||| || || || |||| ||| | || |||||||| ||| |||| ||||| ||||
Db       6823 TTGATGCAATCATAGCAGAACACTTCAAGCAAGGCGACCCGGTACTGGAGACGGATATCG 6882

Qy       1720 CGTCGTTTGATAAAAGCGAAGACGACGCTATCGCCATTTCGGCGTTGATGATCCTTGAGG 1779
              | || || || |||||| |||||||||||||| ||  |  | |  |||||||| | ||||
Db       6883 CATCATTCGACAAAAGCCAAGACGACGCTATGGCGTTAACCGGTCTGATGATCTTGGAGG 6942

Qy       1780 ACTTAGGTGTCGACCAACCGCTCTTAGATTTGATAGAGGCGGCGTTCGGCAATATCACAT 1839
              || | ||||| || ||||| ||  | || ||||| |||   || || ||  | ||  |||
Db       6943 ACCTGGGTGTGGATCAACCACTACTCGACTTGATCGAGTGCGCCTTTGGAGAAATATCAT 7002

Qy       1840 CTGTGCACCTACCTACAGGAACGAGGTTTAAATTTGGTGCCATGATGAAATCCGGTATGT 1899
```


Db 7003 CCACCCATCTACCTACGGGTACTCGTTTTAAATTCGGGGCGATGATGAAATCCGGAATGT 7062
Qy 1900 TCTTAACGCTGTTTGTCAACACACTAGTCAATATCATGATTGCTAGCAGAGTACTACGTG 1959
Db 7063 TCCTCACACTTTTTGTCAACACAGTTTTGAATGTCGTTATCGCCAGCAGAGTACTAGAAG 7122
Qy 1960 AACGGTTAACCACGTCAGCGTGCGCGGCCTCTATCGGCGACGATAACATAGTGCATGGTG 2019
Db 7123 AGCGGCTTAAAACGTCCAGATGTGCAGCGTTCATTGGCGACGACAACATCATACATGGAG 7182
Qy 2020 TCGTCTCCGACACCTTGATGGCGGAGAGATGCGCCACTTGGCTGAACATGGAAGTAAAAA 2079
Db 7183 TAGTATCTGACAAAGAAATGGCTGAGAGGTGCGCCACCTGGCTCAACATGGAGGTTAAGA 7242
Qy 2080 TTATTGATGCAGTTATTGGTATCAAAGCACCCTACTTCTGTGGGGGATTTATCCTGGTGG 2139
Db 7243 TCATCGACGCAGTCATCGGTGAGAGACCACCTTACTTCTGCGGCGGATTTATCTTGCAAG 7302
Qy 2140 ACCAGATAACAGGCACAGCCTGCAGAGTCGCAGACCCTCTAAAAAGGCTTTTTAAGCTTG 2199
Db 7303 ATTCGGTTACTTCCACAGCGTGCCGCGTGGCGGATCCCCTGAAAAGGCTGTTTAAGTTGG 7362
Qy 2200 GAAAACCATTGCCAGTCGATGATACCCAAGACTGCGACCGCCGCCGGGCACTGCATGATG 2259
Db 7363 GTAAACCGCTCCCAGCCGACGACGAGCAAGACGAAGACAGAAGACGCGCTCTGCTAGATG 7422
Qy 2260 AAGCAATGCGATGGAACAGAATTGGAATTACGGACGAGTTAGTGAAGGCCGTAGAATCCA 2319
Db 7423 AAACAAAGGCGTGGTTTAGAGTAGGTATAACAGGCACTTTAGCAGTGGCCGTGACGACCC 7482
Qy 2320 GATACGAGATCATACTGGCAGGCCTGATCATCACGTCTCTGTCCACGTTAGCCGAAAGCG 2379
Db 7483 GGTATGAGGTAGACAATATTACACCTGTCCTACTGGCATTGAGAACTTTTGCCCAGAGCA 7542
Qy 2380 TTAAGAACTTCAAGAGCATAAGAGGGAGCCCAATCACCCTCTACGGCTGACCTAAATAGG 2439
Db 7543 AAAGAGCATTCCAAGCCATCAGAGGGGAAATAAAGCATCTCTACGGTGGTCCTAAATAGT 7602
Qy 2440 TGACGTAGTAGACACGCACCTAC----------CCACCGGCAGAATGTTTCCATACC 2486
Db 7603 CAGCATAGTACATTTCATCTGACTAATACTACAACACCACCACCATGAATAGAGGATTCT 7662
Qy 2487 CTCAGCTGAACTTTCCACCAGTTTACCCTACAAATCCGATGGCTTACCGAGATCCAAACC 2546
Db 7663 TTAACATGCTCGGCCGCCGCCCCTTCCCGGCCCCCACTGCCATGTGGAGGCCGCGGAGAA 7722
Qy 2547 CTCCTAGGCGCCGCTGGAGGCCGTTTCGGCCCCCGCTGGCTGCTCAAATCGAAGATCTTA 2606
Db 7723 GGAGGCAGGCGGCCCCGATGCCTGCCCGCAACGGGCTGGCTTCTCAAATCCAGCAACTGA 7782
Qy 2607 GGAGGTCGATAGTCAACTTGACTTTCAAACAACGATCACCTAATCCGCCGCCAGGTCCA- 2665
Db 7783 CCACAGCCGTCAGTGCCCTAGTCATTGGACAGGCAACTAGACCTCAACCCCCACGTCCAC 7842
Qy 2666 -----CCGCCAAAGAAGAAGAAGAGTGCTCCTAAGCCAAAACCTACTCAGCCTAAAAAGA 2720
Db 7843 GCCCGCCACCGCGCCAGAAGAAGCAGGCGCCCAAGCAACCACCGAAGCCGAAGAAACCAA 7902
Qy 2721 AGAAGCAGCAAGCCAAGAGGACGAAAC---GCAAGCCTAAACCAGGGAAACGACAACGTA 2777
Db 7903 AAACGCAGGAGAAGAAGAAGAAGCAACCTGCAAAACCCAAACCCGGAAAGAGACAGCGCA 7962
Qy 2778 TGTGTATGAAGTTGGAGTCGGACAAGACATTTCCGATCATG---CTGAACGGCCAAGTGA 2834
Db 7963 TGGCACTTAAGTTGGAGGCCGACAGATTGTTCGACGTCAAGAACGAGGACGGAGATGTCA 8022
Qy 2835 ATGGATATGCCTGCGTTGTCGGAGGAAGGCTGATGAAACCACTCCACGTTGAAGGAAAAA 2894
Db 8023 TCGGGCACGCACTGGCCATGGAAGGAAAGGTAATGAAACCTCTGCACGTGAAAGGAACCA 8082
Qy 2895 TTGATAATGAGCAATTAGCGGCCGTGAAATTGAAGAAGGCTAGCATGTACGACTTGGAGT 2954
Db 8083 TCGACCACCCTGTGCTATCAAAGCTCAAATTTACCAAGTCGTCAGCATACGACATGGAGT 8142
Qy 2955 ACGGCGACGTTCCCCAGAACATGAAATCAGACACGCTGCAGTACACCAGCGACAAACCAC 3014
Db 8143 TCGCACAGTTGCCAGTCAACATGAGAAGTGAGGCATTCACCTACACCAGTGAACACCCCG 8202
Qy 3015 CGGGCTTCTACAACTGGCACCACGGCGCAGTCCAGTATGAGAATGGGAGATTTACCGTAC 3074
Db 8203 AAGGATTCTATAACTGGCACCACGGAGCGGTGCAGTATAGTGGAGGTAGATTTACCATCC 8262
Qy 3075 CGAGAGGAGTGGGCGGGAAAGGCGACAGCGGAAGACCGATCCTGGACAACAGAGGCAGAG 3134
Db 8263 CTCGCGGAGTAGGAGGCAGAGGAGACAGCGGTCGTCCGATCATGGATAACTCCGGTCGGG 8322
Qy 3135 TTGTGGCTATTGTTCTAGGAGGTGCAAATGAGGGCACGCGTACGGCGCTTTCAGTGGTCA 3194
Db 8323 TTGTCGCGATAGTCCTCGGTGGCGCTGATGAAGGAACACGAACTGCCCTTTCGGTCGTCA 8382
Qy 3195 CTTGGAACCAGAAAGGGGTGACCATTAGGGATACCCCCGAAGGTTCTGAACCGTGGT--- 3251
Db 8383 CCTGGAATAGTAAAGGGAAGACAATTAAGACGACCCCGGAAGGGACAGAAGAGTGGTCCG 8442
Qy 3252 ------CACTAGTTACAGCGCTATGCGTGCTTTCGAATGTCACGTTCCCATGCGACAAAC 3305
Db 8443 CAGCACCACTGGTCACGGCAATGTGTTTGCTCGGAAATGTGAGCTTCCCATGCGACCGCC 8502
Qy 3306 CACCCGTGTGCTATTCACTGACGCCAGAACGAACACTCGACGTGCTCGAAGAGAACGTCG 3365
Db 8503 CGCCCACATGCTATACCCGCGAACCTTCCAGAGCCCTCGACATCCTTGAAGAGAACGTGA 8562
Qy 3366 ACAATCCAAATTACGACACGCTGCTGGAGAACGTCTTGAAATGTCCATCAC---GCCGGC 3422
Db 8563 ACCATGAGGCCTACGATACCCTGCTCAATGCCATATTGCGGTGCGGATCGTCTGGCAGAA 8622
Qy 3423 CCAAACGAAGCATTACCGATGACTTCACACTGACCAGTCCCTACCTGGGGTTCTGCCCGT 3482
Db 8623 GCAAAAGAAGCGTCATTGACGACTTTACCCTGACCAGCCCCTACTTGGGCACATGCTCGT 8682
Qy 3483 ATTGCAGACACTCAACGCCGTGTTTCAGCCCAATAAAAATTGAGAACGTGTGGGACGAAT 3542
Db 8683 ACTGCCACCATACTGTACCGTGCTTCAGCCCTGTTAAGATCGAGCAGGTCTGGGACGAAG 8742
Qy 3543 CTGATGATGGATCGATTAGAATCCAGGTCTCGGCACAATTCGGCTACAATCAGGCAGGCA 3602
Db 8743 CGGACGATAACACCATACGCATACAGACTTCCGCCCAGTTTGGATACGACCAAAGCGGAG 8802
Qy 3603 CTGCGGATGTCACCAAATTCCGTTACATGTCTTTCGACCACGACCATGACATCAAGGAAG 3662
Db 8803 CAGCAAGCGCAAACAAGTACCGCTACATGTCGCTTAAGCAGGATCACACCGTTAAAGAAG 8862
Qy 3663 ACAGTATGGAGAAAATAGCTATCAGCACATCTGGACCCTGCCGTCGTCTTGGCCACAAAG 3722
Db 8863 GCACCATGGATGACATCAAGATTAGCACCTCAGGACCGTGTAGAAGGCTTAGCTACAAAG 8922
Qy 3723 GGTACTTCCTGTTAGCTCAATGTCCTCCAGGTGACAGTGTAACCGTCAGTATCACGAGCG 3782
Db 8923 GATACTTTCTCCTCGCAAAATGCCCTCCAGGGGACAGCGTAACGGTTAGCATAGTGAGTA 8982
Qy 3783 GAGCATCTGAGAATTCATGCACCGTGGAGAAAAAGATCAGGAGGAAGTTTGTCGGTAGAG 3842
Db 8983 GCAACTCAGCAACGTCATGTACACTGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGG 9042
Qy 3843 AGGAGTACTTGTTCCCACCCGTCCATGGAAAGCTGGTAAAGTGCCACGTTTACGATCACT 3902
Db 9043 AAAAATATGATCTACCTCCCGTTCACGGTAAAAAAATTCCTTGCACAGTGTACGACCGTC 9102
Qy 3903 TGAAGGAGACGTCTGCCGGGTACATAACCATGCACAGGCCAGGCCCACACGCGTATAAGT 3962
Db 9103 TGAAAGAAACAACTGCAGGCTACATCACTATGCACAGGCCGAGACCGCACGCTTATACAT 9162
Qy 3963 CCTATCTGGAGGAAGCGTCAGGCGAAGTGTACATTAAACCACCTTCTGGCAAGAACGTCA 4022
Db 9163 CCTACCTGGAAGAATCATCAGGGAAAGTTTACGCAAAGCCGCCATCTGGGAAGAACATTA 9222


Qy 4023 CCTACGAATGTAAGTGTGGCGACTACAGCACAGGTATCGTGAGCACGCGAACGAAGATGA 4082
Db 9223 CGTATGAGTGCAAGTGCGGCGACTACAAGACCGGAACCGTTTCGACCCGCACCGAAATCA 9282
Qy 4083 ACGGCTGCACTAAAGCAAAACAGTGCATTGCCTACAAGAGCGACCAAACGAAATGGGTCT 4142
Db 9283 CTGGTTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAACGAAGTGGGTCT 9342
Qy 4143 TCAACTCGCCGGATCTTATTAGGCACACAGACCACTCAGTGCAAGGTAAATTGCACATTC 4202
Db 9343 TCAACTCACCGGACTTGATCAGACATGACGACCACACGGCCCAAGGGAAATTGCATTTGC 9402
Qy 4203 CATTCCGCTTGACACCGACAGTCTGCCCGGTTCCGTTAGCTCACACGCCTACAGTCACGA 4262
Db 9403 CTTTCAAGTTGATCCCGAGTACCTGCATGGTCCCTGTTGCCCACGCGCCGAATGTAATAC 9462
Qy 4263 AGTGGTTCAAAGGCATCACCCTCCACCTGACTGCAATGCGACCAACATTGCTGACAACGA 4322
Db 9463 ATGGCTTTAAACACATCAGCCTCCAATTAGATACAGACCACTTGACATTGCTCACCACCA 9522
Qy 4323 GAAAATTGGGGCTGCGAGCAGACGCAACAGCAGAATGGATTACAGGGTCTACATCCAGGA 4382
Db 9523 GGAGACTAGGGGCAAACCCGGAACCAACCACTGAATGGATCGTCGGAAAGACGGTCAGAA 9582
Qy 4383 ATTTTTCTGTGGGGCGAGAAGGGCTGGAGTACGTATGGGGTAACCATGAACCAGTCAGAG 4442
Db 9583 ACTTCACCGTCGACCGAGATGGCCTGGAATACATATGGGGAAATCATGAGCCAGTGAGGG 9642
Qy 4443 TCTGGGCCCAGGAGTCGGCACCAGGCGACCCACATGGATGGCCGCATGAGATCATCATCC 4502
Db 9643 TCTATGCCCAAGAGTCAGCACCAGGAGACCCTCACGGATGGCCACACGAAATAGTACAGC 9702
Qy 4503 ACTATTATCATCGGCATCCAGTCTACACTGTCATTGTGCTGTGTGGTGTCGCTCTTGCTA 4562
Db 9703 ATTACTACCATCGCCATCCTGTGTACACCATCTTAGCCGTCGCATCAGCTACCGTGGCGA 9762
Qy 4563 TCCTGGTAGGCACTGCATCATCAGCAGCTTGCATCGCCAAAGCAAGAAGAGACTGCCTGA 4622
Db 9763 TGATGATTGGCGTAACTGTTGCAGTGTTATGTGCCTGTAAAGCGCGCCGTGAGTGCCTGA 9822
Qy 4623 CGCCATACGCGCTTGCACCGAACGCAACGGTACCCACAGCATTAGCGGTTTTGTGCTGCA 4682
Db 9823 CGCCATACGCCCTGGCCCCAAACGCCGTAATCCCAACTTCGCTGGCACTCTTGTGCTGCG 9882
Qy 4683 TTCGGCCAACCAACGCTGAAACATTTGGAGAAACTTTGAACCATCTGTGGTTTAACAACC 4742
Db 9883 TTAGGTCGGCCAATGCTGAAACGTTCACCGAGACCATGAGTTACTTGTGGTCGAACAGTC 9942
Qy 4743 AACCGTTTCTCTGGGCACAGTTGTGCATTCCTCTGGCAGCGCTTGTTATTCTGTTCCGCT 4802
Db 9943 AGCCGTTCTTCTGGGTCCAGTTGTGCATACCTTTGGCCGCTTTCATCGTTCTAATGCGCT 10002
Qy 4803 GCTTTTCATGCTGCATGCCTTTTTTATTGGTTGCAGGCGTCTGCCTGGGGAAGGTAGACG 4862
Db 10003 GCTGCTCCTGCTGCCTGCCTTTTTTAGTGGTTGCCGGCGCCTACCTGGCGAAGGTAGACG 10062
Qy 4863 CCTTCGAACATGCGACCACTGTGCCAAATGTTCCGGGGATCCCGTATAAGGCGTTGGTCG 4922
Db 10063 CCTACGAACATGCGACCACTGTTCCAAATGTGCCACAGATACCGTATAAGGCACTTGTTG 10122
Qy 4923 AACGCGCAGGTTACGCGCCACTTAACCTGGAGATCACGGTCGTCTCATCGGAATTAACAC 4982
Db 10123 AAAGGGCAGGGTATGCCCCGCTCAATTTGGAGATCACTGTCATGTCCTCGGAGGTTTTGC 10182
Qy 4983 CTTCAACTAACAAGGAGTACGTGACCTGCAAATTCCACACAGTCATTCCTTCACCACAAG 5042
Db 10183 CTTCCACCAACCAAGAGTACATTACCTGCAAATTCACCACTGTGGTCCCCTCCCCAAAAA 10242
Qy 5043 TTAAATGCTGCGGGTCCCTCGAGTGCAAGGCATCCTCAAAGGCGGATTACACATGCCGCG 5102
Db 10243 TCAAATGCTGCGGCTCCTTGGAATGTCAGCCGGCCGCTCATGCAGACTATACCTGCAAGG 10302
Qy 5103 TTTTTGGCGGTGTGTACCCTTTCATGTGGGGAGGCGCACAATGCTTCTGTGACAGTGAGA 5162
Db 10303 TCTTCGGAGGGGTCTACCCCTTTATGTGGGGAGGAGCGCAATGTTTTTGCGACAGTGAGA 10362
Qy 5163 ACACACAACTGAGTGAGGCGTACGTCGAGTTCGCTCCAGACTGCACTATAGATCACGCAG 5222
Db 10363 ACAGCCAGATGAGTGAGGCGTACGTCGAATTGTCAGCAGATTGCGCGTCTGACCACGCGC 10422
Qy 5223 TCGCACTAAAAGTTCACACAGCTGCTCTGAAAGTCGGCCTGCGTATAGTATACGGCAACA 5282
Db 10423 AGGCGATTAAGGTGCACACTGCCGCGATGAAAGTAGGACTGCGTATTGTGTACGGGAACA 10482
Qy 5283 CCACCGCGCACCTGGATACGTTTGTCAATGGCGTCACGCCAGGTTCCTCACGGGACCTGA 5342
Db 10483 CTACCAGTTTCCTAGATGTGTACGTGAACGGAGTCACACCAGGAACGTCTAAAGACTTGA 10542
Qy 5343 AGGTCATAGCAGGGCCGATATCAGCCGCTTTTTCACCCTTTGACCATAAGGTCGTCATCA 5402
Db 10543 AAGTCATAGCTGGACCAATTTCAGCATCGTTTACGCCATTCGATCATAAGGTCGTTATCC 10602
Qy 5403 GAAAGGGGCTTGTTTACAACTACGACTTCCCTGAGTATGGAGCTATGAAACCAGGAGCGT 5462
Db 10603 ATCGCGGCCTGGTGTACAACTATGACTTCCCGGAATATGGAGCGATGAAACCAGGAGCGT 10662
Qy 5463 TCGGCGATATTCAAGCATCCTCGCTTGATGCTACAGACATAGTAGCCCGCACTGACATAC 5522
Db 10663 TTGGAGACATTCAAGCTACCTCCTTGACTAGCAAGGATCTCATCGCCAGCACAGACATTA 10722
Qy 5523 GGCTGCTGAAGCCTTCTGTCAAGAACATCCACGTCCCCTACACCCAAGCAGTATCAGGGT 5582
Db 10723 GGCTACTCAAGCCTTCCGCCAAGAACGTGCATGTCCCGTACACGCAGGCCTCATCAGGAT 10782
Qy 5583 ATGAAATGTGGAAGAACAACTCAGGACGACCCCTGCAAGAAACAGCACCATTTGGATGTA 5642
Db 10783 TTGAGATGTGGAAAAACAACTCAGGCCGCCCACTGCAGGAAACCGCACCTTTCGGGTGTA 10842
Qy 5643 AAATTGAAGTGGAGCCTCTGCGAGCGTCTAACTGTGCTTACGGGCACATCCCTATCTCGA 5702
Db 10843 AGATTGCAGTAAATCCGCTCCGAGCGGTGGACTGTTCATACGGGAACATTCCCATTTCTA 10902
Qy 5703 TTGACATCCCTGATGCAGCTTTTGTGAGATCATCAGAATCACCAACAATTTTAGAAGTTA 5762
Db 10903 TTGACATCCCGAACGCTGCCTTTATCAGGACATCAGATGCACCACTGGTCTCAACAGTCA 10962
Qy 5763 GCTGCACAGTAGCAGACTGCATTTATTCTGCAGACTTTGGTGGTTCTCTAACATTACAGT 5822
Db 10963 AATGTGAAGTCAGTGAGTGCACTTATTCAGCAGACTTCGGCGGGATGGCCACCCTGCAGT 11022
Qy 5823 ACAAAGCTGACAGGGAGGGACATTGTCCAGTTCACTCCCACTCCACGACAGCTGTTTTGA 5882
Db 11023 ATGTATCCGACCGCGAAGGTCAATGCCCCGTACATTCGCATTCGAGCACAGCAACTCTCC 11082
Qy 5883 AGGAAGCGACCACACATGTGACTGCCGTAGGCAGCATAACACTACATTTTAGCACATCGA 5942
Db 11083 AAGAGTCGACAGTACATGTCCTGGAGAAAGGAGCGGTGACAGTACACTTTAGCACCGCGA 11142
Qy 5943 GCCCACAAGCAAATTTTATAGTTTCGCTATGCGGCAAGAAGTCCACCTGCAATGCTGAAT 6002
Db 11143 GTCCACAGGCGAACTTTATCGTATCGCTGTGTGGGAAGAAGACAACATGCAATGCAGAAT 11202
Qy 6003 GTAAACCACCGGCCGACCACATAATTGGAGAACCACATAAAGTCGACCAAGAATTCCAGG 6062
Db 11203 GTAAACCACCAGCTGACCATATCGTGAGCACCCCGCACAAAAATGACCAAGAATTTCAAG 11262
Qy 6063 CGGCAGTTTCCAAAACATCTTGGAACTGGCTGCTTGCACTGTTTGGGGGAGCATCATCCC 6122
Db 11263 CCGCCATCTCAAAAACATCATGGAGTTGGCTGTTTGCCCTTTTCGGCGGCGCCTCGTCGC 11322
Qy 6123 TCATTGTTGTAGGACTTATAGTGTTGGTCTGCAGCTCTATGCTTATAAACACACGTAGAT 6182
Db 11323 TATTAATTATAGGACTTATGATTTTTGCTTGCAGCATGATGCTGACTAGCACACGAAGAT 11382
Qy 6183 GACTGAGCGCGGACACTGA-CATAGCGGTAAAACTCGATGTACTTCCGAGGAAGCGTGGT 6241

```
            ||| |       || ||| |  | | | ||||||||||||||||||||||  |  ||
Db      11383 GACCGCTACGCCCCAATGATCCGACCAGCAAAACTCGATGTACTTCCGAGGAACTGATGT 11442

Qy       6242 GCATAATGCCACGCGCCGCT 6261
              |||||||||  |  || | |
Db      11443 GCATAATGCATCAGGCTGGT 11462
```

Search completed: November 16, 2003, 09:06:06
Job time : 380 secs